A GENERATIVE MODEL OF TONAL TENSION

AND ITS APPLICATION IN

DYNAMIC REALTIME SONIFICATION

A Thesis
Presented to
The Academic Faculty

by

Ryan Nikolaidis

In Partial Fulfillment
of the Requirements for the Degree
Master of Science in Music Technology in the
School of Department of Music

Georgia Institute of Technology
December 2011

A GENERATIVE MODEL OF TONAL TENSION

AND ITS APPLICATION IN

DYNAMIC REALTIME SONIFICATION

Approved by:

Dr. Gil Weinberg, Advisor
Department of Music
Georgia Institute of Technology

Dr. Jason Freeman
School of Department of Music
Georgia Institute of Technology

Dr. Parag Chordia
School of Department of Music
Georgia Institute of Technology

Date Approved: June 2011

To my loving and supportive wife, Jenna

ACKNOWLEDGEMENTS

I wish to thank Gil Weinberg, director of the Center for Music Technology, with whose guidance and support made my work possible.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
SUMMARY	x
CHAPTER	
1 Introduction	1
1.1 Overview	1
1.2 Contribution	2
2 Related Work	5
2.1 Perception and Cognition	5
2.2 Generative Music	21
2.3 Sonification	25
3 Accessible Aquarium	29
4 Hypothesis	33
5 Methods	35
6 Implementation	37

6.1 Tracking Visual Tension 37

6.2 Rhythm Generation 38

6.3 Melody Generation 40

6.4 Harmony Generation 43

7 Evaluation 50

7.1 Study 1: Melodic Attraction and Rhythmic Stability 50

7.2 Study 2: Sonification 55

8 Conclusions 60

9 Future Work 61

APPENDIX A: Chord Transition Table 62

REFERENCES 64

LIST OF TABLES

Page

Table 1: Anchoring strength table 10

Table 2: Results showing mean, standard deviation, and drift 17

Table 3: Results from stimuli comparison and standard 19

Table 4: Results and analysis of mean mentioning of features 31

Table 5: Sonification mappings table 37

Table 6: Relative note distance 41

LIST OF FIGURES

Page

Figure 1: Helical representation of pitch perception 7

Figure 2: Shepard's double helix model of pitch distance 8

Figure 3: Geometric representation of pitch relatedness 9

Figure 4: Subject response time 14

Figure 5: Anchoring Space 15

Figure 6: Divergence from the stimuli duration and the mean produced ratio 17

Figure 7: Distribution based on the proportion of all categories chosen 21

Figure 8: The interaction between rhythm-, pitch-, and harmony- generation modules 35

Figure 9: Rhythm Generation 39

Figure 10: Mapping speed to note density 40

Figure 11: Harmony Generation Diagram 44

Figure 12: Lerdahl's model of the theoretical distance between two chords 45

Figure 13: Example of calculating distance between chords 45

Figure 14: Results from Lerdahl's evaluation of his Tension model 46

Figure 15: GMR in perceived tension as compared to change in register. 52

Figure 16: GMR in perceived tension as compared to change in instrument 53

Figure 17: GMR in perceived tension as compared to change in the input tension 54

Figure 18: ANOVA analysis of responses to statement 1 57

Figure 19: ANOVA analysis of responses to statement 2 58

Figure 20: ANOVA analysis of responses to statement 3 58

Figure 21: ANOVA analysis of responses to statement 4 59

SUMMARY

This thesis presents the design and implementation of a generative model of tonal tension. It further describes the application of the generative model in realtime sonification. The thesis discusses related theoretical work in musical fields including generative system design, sonification, and perception and cognition. It highlights a review of the related research from historical to contemporary work. It contextualizes this work in informing the design and application of the generative model of tonal tension. The thesis concludes by presenting a formal evaluation of the system. The evaluation consists of two independent subject-response studies assessing the effectiveness of the generative system to create tonal tension and map it to visual parameters in sonification.

CHAPTER 1

INTRODUCTION

This thesis describes my generative model of tonal tension. The model generates multiple monophonic lines of music with respect to an input tension parameter. It contextualizes the work with its application in the realtime sonification of fish movements.

Music is multidimensional. We often define it in dimensions including rhythm, loudness, instrumentation, and emotion. Tension is one of these dimensions. It relies on the sense of expectation (Meyer, 1967). Theorist Leonard Meyer explains that music can invoke an expectation for future events, and the creation and suspension of this expectation yields musical tension. Likewise, music can fulfill these expectations, releasing tension and creating a sense of stability and relaxation. In a broad sense, any aspect of music capable of creating or fulfilling expectation (e.g. rhythm, volume, timbre...etc.) affects musical tension (Farbood, 2001); an idea that lead musicologist Arnold Schering to describe music as “nothing other than a continuous balancing of feuding sonic and dynamic principles” (Christensen, 2001). This work is particularly focused on the perceptual response to tonal tension, musical tension created by changes in the organization of pitches (Lerdahl, 2001).

1.1 Overview

This research concerns the design and implementation of the real-time music generation algorithm, it's application in sonification of a visual display, and a user studies conducted to evaluate the effectiveness of the approach with human listeners. In recent years a large body of research in music perception has focused on modeling aspects of tension in music. These models aim to predict how we hear tension based on some analysis of certain musical features. My work combines the results of some of the more prominent and proven models in order to generate, rather than analyze, music governed by an expected level of tension (which I refer to as the input tension level). With respect to sonification, I explore how to map this generative tension model to the visual tracking of fish location in sonification research titled the Accessible Aquarium project. Furthermore, the thesis will address a dichotomy in sonification research, presenting research that either uses auditory feedback to convey information or focuses more on the resulting musical aesthetic.

The thesis begins by laying out the goals and challenges addressed throughout the project, and continues by describing the related work. It details previous attempts at musical sonification systems implemented in conjunction with the Accessible Aquarium project, providing a background for which to distinguish my distinct contribution to the project. Then the thesis discusses the project's design with generative features ultimately directed by the manipulation of tonal tension. It then presents technical choices, such as the use of Fred Lerdahl's formulas for analysis of tension in melodic and harmonic aspects of music (Lerdahl 2001) as a model for a generative application of these parameters. The thesis shows the correlation between the generative engine and cognitive theory and details the incorporation of input variables as facilitators of low- and high-level mappings of visual information. It concludes with a description of two user studies as well as self-evaluation of the work, and discusses prospective future work including improvements to our current modeling method and developments in additional high-level precepts.

1.2 Contribution

The work advances research in music perception, generative systems, and sonification. In music perception, it modifies and combines the results of previous models into a single comprehensive model, in which the accuracy is evaluated in a series of subject-response studies. Rather than using the comprehensive model of tension for analysis, the work uses a predictive model in order to generate music of reliably variable perceived tension. Finally, the research addresses an overlooked yet ongoing split in sonification work, and offers a solution through simultaneous high- and low-level mappings in order to convey non-musical information while attending to musical aesthetic.

While several theories try to explain how and why we hear tension as the result of certain features in music, none model it generatively. For instance, Fred Lerdahl's Model of tonal tension focuses on how analyzing the organization of pitch material influences the perception of tension. Similarly, Desain and Honings work addresses specifically the perception of rhythmic stability. In designing a generative system based on tension I consider the influence of tension across all of these categories and, in doing so, develop a more comprehensive model of tension that takes multiple features into consideration.

Additionally, the body of existing sonification research has not yet successfully addressed the need for a method to bridge the divide between information conveying and musically engaging implementations. Scientists and researchers in sonification often develop tools to explore visual or non-sonic information in an auditory domain or to aid the visually impaired in everyday tasks. The results from these systems often involve simple mappings of low-level features, resulting in uninteresting music. For example, heart-rate monitors and Geiger counters use nothing more than changes in beep frequency to represent heart-rate and radiation levels, respectively. The corollary includes artists that want to create music driven by some non-musical stimuli, with the end-result being the generation of some interesting musical output with little direct conveyance of the non-musical stimuli. For instance, the Reeds project utilizes an array of weather data from multiple stations as the seed material for a “synthesis instrument”, producing music that lacks any meaningful representation of the original data. However, some recent researchers and artists have begun exploring this continuum of data representation and musical aesthetic. In “Interactive Sonification of Neural Activity”, Weinberg and Thatcher directly address research split at both ends of this continuum, leading to the focus of their sonification work BrainWaves to “provide an aesthetically satisfying and educationally useful representation of complex datasets of neural activity” (Weinberg and Thatcher, 2006). Ben-Tal and Berger also detail the dichotomy, dividing the focus of “Creative Aspects of Sonification” between the data representation and the “purely aesthetic use of sonified statistical data” approaches (Ben-Tal and Berger, 2004). In her sonification of atmospherics and weather, Andrea Polli approaches mapping the data to music with the intention of creating musical compositions (Polli, 2004). She focuses on exploring different combinations of direct (one-to-one) low-level mappings of data to musical features that generate her desired musical results. For instance, by mapping wind speed directly to the amplitude of generated sounds, she explains that it creates “sonic activity and excitement”. She elaborates by explaining the issue of variability in these atmospheric features based on elevation, which she resolves through a thoughtful application of global and local scaling (scaling respectively invariant or exclusive to elevation). Global scaling for wind speed to sound amplitude created “...the compositions building and receding in intensity”. Whereas features such as pitch and filtering coefficients, need local scaling in order to account for the dramatic variability between elevations. I believe that by combining the design of an effective and engaging generative music system with the sonification of low- and high- level musical and visual features, both information conveying and musically engaging criteria can simultaneously be achieved.

CHAPTER 2

RELATED WORK

The entirety of the research aims to build on theoretical models of tension in order to generate music that sonifies some non-musical stimuli. As such, it draws on a diverse range of related work. In this section I present a narrative of theoretical background and previous similar research that has informed the design in which I based this work

2.1 Perception and Cognition

Designing the model of tonal tension, I divided the focus between three major musical feature spaces pertaining to the organization of notes: melodic attraction, harmonic expectation, and rhythmic stability; all of which have been shown to have a strong affect on the perception of tension (Farbood, 2001). Melodic attraction (Margulis, 2005; Lerdahl, 2001) measures the perceptual tension-resolution created by movements between individual contiguous notes. Harmonic expectation describes the perceived likelihood of one group of simultaneously generated notes moving to the next. Rhythmic stability refers to the organization of notes in time; it defines how different arrangements of notes in time can create more or less tension. Research in all of these areas has led to the comprehensive understanding of tension in music.

2.1.1 Melodic Attraction

In A Generative Theory of Tonal Music (GTTM), Lerdahl and Jackendoff (1996) presented a theoretical approach to detail how we structurally hear tonal music. This work eventually led to Lerdahl's study and theory of tension in tonal music (Lerdahl 2001). Here he formalized a series of equations to model tension, each equation resulting from a unique musical feature. These features included melodic attraction, pitch space (expectancy), and surface dissonance.

Lerdahl's model incorporated the results of several independent studies of these features, which he theorized would affect the perception of tension. Drobisch (1855) and Shepard's (1982) geometrical models proposed a formal relatedness between various pitches of a diatonic system. These models provided a foundation for research in melodic attraction. The ideas developed further through the work of researchers including Bharucha (1996), Krumhansl (1979), Hutchinson and Knopoff (1978) , all of whom influenced the theory of tension proposed by Lerdahl. In defining melodic attraction, Lerdahl built on Bharucha's proposed theory of giving relative priority to scale and chord tones. Lerdahl used models proposed and tested by Krumhansl (1979) and Bharucha (1996) to define relatedness between tones. Finally, he adapted Hutchinson and Knopoff's (1978) model of dissonance into his comprehensive model of tension, explicitly treating dissonance relative to the previous material.

Similar to Lerdahl's work, Narmour introduced a model of melodic expectation; his model looked exclusively at pitch-to-pitch transitions. He proposed that any two contiguous pitches infer a third. A small interval between the first two pitches suggests a third pitch that continuous in the same direct as the interval between the first two. Contrastingly, a large interval between the two pitches implies a third pitch in the reverse direction, the 'gap-fill' phenomenon.

These studies of musical expectation build on earlier work in music perception focused on finding a way to represent pitches that modeled our perception of them. The Chroma Circle represents all twelve pitch-classes as a circle of contiguous pitches that are each a minor second in distance. Explaining the perception of pitch, Drobisch proposed a geometric model (Drobisch 1982). He presented a helical transformation of the Chroma Circle.

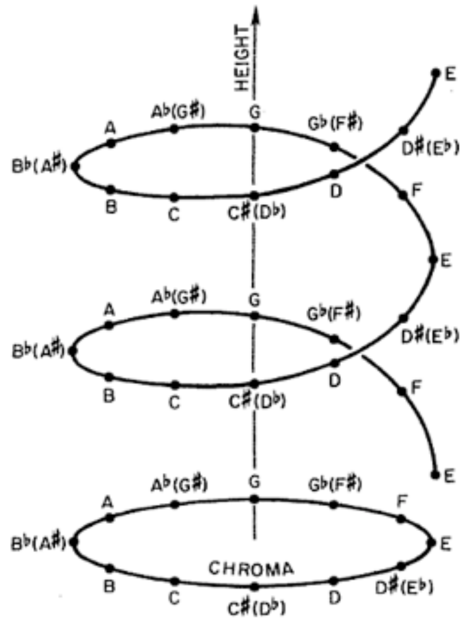

Figure 1. Helical representation of pitch perception (Shepard 1982)

In this transformation, Figure 1, Drobisch extended the Chroma Circle's representation of relative distance in pitch class to also represent change in pitch height. From this new model, a distance relationship could be shown between any two pitches, not just the twelve pitch classes. Extending Drobisch's work, Shepard developed a more sophisticated model. In this model he represented the distance between pitches as a double helix structure.

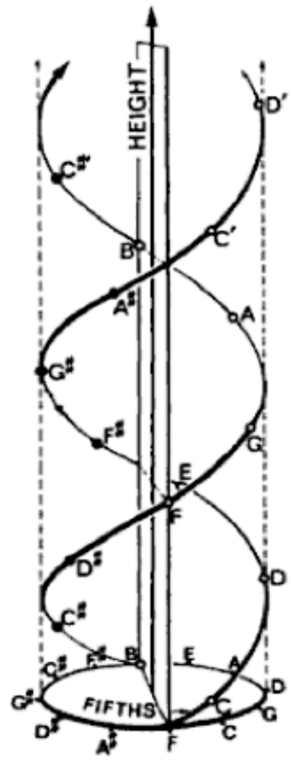

Figure 2. Shepard's double helix model of pitch distance (Shepard 1982)

In designing this geometric model Shepard used two helixes that both consisted of a strand of contiguous whole tone pitches. He constructed each from a different whole tone series. As shown in Figure 2, when projected onto an orthogonal surface the double helix model transforms into a Circle of Fifths, which represents pitch space distance according to key similarity.

Krumhansl conducted several studies investigating the perception of pitch in the context of tonality (Krumhansl 1979). She found that subjects tend to group pitch relatedness according to Chroma (the set of pitches related by octaves, pitch class). In particular, pitches were related in terms of chord and diatonic context. Building on Shepard's geometric approach, through multidimensional scaling she derived a geometric cone-shaped model that demonstrated her relatedness ordering:

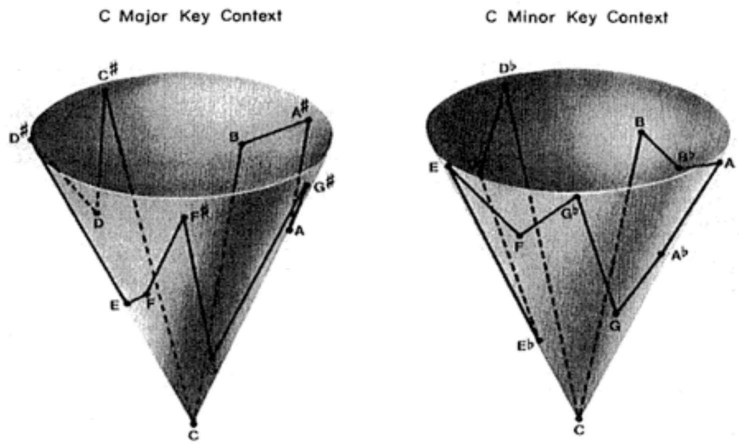

Figure 3. Geometric representation of pitch relatedness (Krumhansl and Kessler 1982)

Figure 3 shows how pitches were related according to modal context. In these geometric representations the distance between pitches represents their relatedness. Additionally, the pitches' distance from the tip of the cone represents their relative stability. Near the vertex of the cone, in each case, the tonic triad was found to be more stable than other pitches, and in particular the root of the scale and triad the most stable.

Building on this, Bharucha proposes a theory of melodic anchoring in which unstable tones are attracted to those more stable and close in pitch (Bharucha 1984). He defines stability in terms of harmonic and diatonic context, in which more stable tones are harmonic and/or diatonic. To test the theory he conducted a subject response study. In the study, subjects listened to a series of melody pairs in which an unstable tone either resolved to a, close in pitch, stable tone or unstable tone. The study affirmed his theory; subjects assigned higher stability ratings to the examples in which the unstable tone resolved to the anchor tone.

Table 1. Anchoring strength table for computing the attraction between pitches.

Strength	Basic Pitch Space (0 = tonic, 11 = leading tone…)										
4	0										
3	0				4			7			
2	0		2		4	5		7		9	
1	0	1	2	3	4	5	6	7	8	9	10

The 0 value in Table 1 represents the root of a scale, 11 its leading tone, and 1 through 10 correspond to the relative ten notes in between. Table 1 demonstrates the anchoring space representing a major tonic triad. As such, the values 0, 4, and 7 have the highest valued anchoring strength because these pitch classes correspond to the harmonic tones of the major tonic triad. In particular, the root of the chord has the highest anchoring rating of all; in Bharucha's theory of attraction, the root is the most stable tone.

Later, in Melodic Anchoring, Bharucha investigated melodic attraction through his neural network model (Bharucha 1996). This self-organizing model connected pitches to chords and keys. Through study and application of real music, he found statistical relationships between harmonic and non-harmonic tones and related this to stability. He was also able to support his earlier theories about the balance between stable and unstable tones .

Building on precepts designed by Bharucha, another perceptual design models musical forces affecting pitch contour after natural forces described by physics (Larson 1997). In his work, Larson details three main contributions as Gravity, Magnetism, and Inertia. Gravity refers to the tendency of unstable pitches to descend towards more stable ones. Similarly, Magnetism describes stronger attraction between closer pitches. Finally, Inertia represents a contour's tendency to continue in a 'same' manner, i.e. if it was descending it is more likely to continue descent rather than change direction.

David Huron (2001) further validates Larson's work in Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles. In the article Huron argues that the aesthetic principals of voice-leading, in common-practice tonality, originate from concepts in music perception. In developing the theory he points to the voice-leading phenomenon of asymmetrical embellishments which he suggests could be attributed to the related work presented above: "...embellishments ... exhibit systematic asymmetries, such as the penchant for resolution by step, but approach by either step or leap. A plausible cognitive answer may be found in the anchoring of unstable tones to stable ones, as demonstrated by Krumhansl (1979) and investigated more thoroughly by Bharucha (1984)".

To model melodic attraction, Lerdahl (Lerdahl 2001; Lerdahl and Krumhansl 2007) formalizes the precepts proposed by Bharucha and Larson. His equation, validated in the studies presented in Tonal Pitch Space, consists of two major components: anchoring strength and relative distance.

$$S=\left(\frac{a_2}{a_1}\right)\left(\frac{1}{n^2}\right)$$

Equation 1. Lerdahl's voice leading stability equation (2001). a1 and a2 represent the previous and next note anchoring strength respectively and n represents the relative step-size from the previous pitch to the next pitch

Lerdahl's model (2001), Equation 1, consists of two components: relative anchoring strength and pitch distance. Reflecting Larson's work, Lerdahl uses the inverse-square of pitch distance in order to parallel Newton's Law of Universal Gravitation, which uses the inverse-square of the radius of a massive object to determine gravitational force. As such, a higher attraction exists between closer pitches. The anchoring strength factor accounts for attraction between less stable to more stable pitches. For instance, in the key of C Major, pitch B is highly attracted to C. However, due to anchoring effects, Pitch C is less attracted to B.

2.1.2 Harmonic Expectation

My algorithm generatively implements and combines two distinct theoretical models of harmonic expectation. Theories of Western Tonal Music often attempt to define harmonic changes in terms of heuristics that model our expectations. Justus and Bharucha (2001) classify these expectations as the result of schemata and veridical information . They use the term schemata to refer to harmonic constructs learned through past experiences and veridical information as familiarity through recent musical memories. Bharucha (1991) proposes that we learn schemata passively, through perceptual learning . As such, by listening to music, we implicitly learn probabilities governing harmonic sequential transitions. Modeling harmonic expectations, theorists such as Fred Lerdahl (2001) and François Pachet (1999) relied heavily on the work of previous perceptual studies. Their work demonstrates two major approaches in expectation modeling. Lerdahl codifies harmonic expectation with a set of heuristics addressing the relatedness of two chords. Contrastingly, Pachet more directly relies on schemata, modeling expectation based on chord transition probability.

Laying the groundwork for these models, Bharucha and Stoeckig (1987) designed and implemented a series of perceptual studies to explore expectancy associated with harmonic transitions. Working under the assumption that harmonic recognition is a modular process they evaluated expectation as the product of processing speed; they propose that more likely events would process faster than that of less likely events. As a result, their studies shared a common design; they presented subjects with a series of more or less expected chords and had them perform some orthogonal task, such as detecting changes in timbre. By measuring the relative changes in response time, they could map the relative expectancy of harmonic transitions. They found that expectancy towards harmonic transitions closely mapped distances represented by the Circle of Fifths.

Similar to this study, Tillman et al. (2003) developed a series of priming studies to test the correlation between chord function and expectation. They presented subjects with a series of chords that ended on a target chord, which varied in function (being either tonic or subdominant). Upon hearing the target chord, subjects were asked to respond to an orthogonal task. Similar to the previous study, they employed measurement of response time as means of evaluating harmonic expectancy. Their findings correlated to stability assumptions in traditional Western music theory; subjects had faster response time for tonic chords, commonly accepted as the functionally most stable chord. In order to further evaluate the slower response time to the subdominant, Tillman et al. conducted an additional study using priming chords that either did or did not clearly establish the key. They found that the slower response time resulted from an 'inhibition' effect, in fact worse in the case where the subdominant target is preceded by a clear establishment of the key. They conclude that the establishment of the key creates expectations not met by the subdominant; sequences without a strongly established key lack these expectations and therefore subject responses do not exhibit the 'inhibition' effect from misdirected expectations.

Tillman et al. (2008) continued the research in a later study in which they compared response time across tonic, subdominant, and dominant target chords

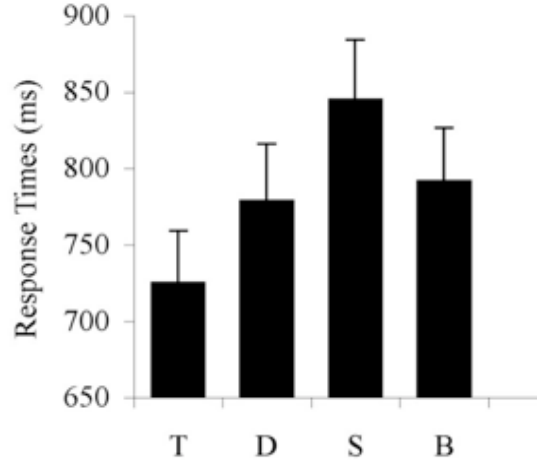

Figure 4. Subject response time with tonic (T), dominant (D), and subdominant (S) targets compared to baseline (B) (Tillman et al., 2008)

They found that response time to the dominant target shorter than the subdominant and longer than the tonic. Comparing these results to a baseline in which none of the chords were related, the dominant chord showed neither a significant cost nor benefit due to expectation. The researchers emphasize that these results support the notion of the dominant chord just second to the tonic in order of importance in Western music theory, and also follow in order of statistical probability (with tonic being the most common).

As the previous perceptual studies showed the connection between chord relatedness and expectation, Lerdahl (2001) modeled a perceived distance measure between chords (determining how related are two chords and therefore the expectation of transition from one to the other). In tonal Pitch Space, Lerdahl proposes a sum of three distinct measurements. These measurements include: diatonic space distance, root distance, and a comparison of anchoring spaces. He describes these anchoring spaces in terms of an implied diatonic space, where anchoring strengths are determined by a combination of key and chord tones. Lerdahl suggests that moving between spaces that are farther apart creates tension. He incorporates two factors in measuring the distance between two spaces. First, he looks at shared tones between the spaces. For example, since A Minor and C Major share the same diatonic space the measurement for this distance is zero. The diatonic spaces implied by C Major and C Minor, however, are three steps apart. Next, root distance refers to the number of movements along the Circle of Fifths between the root of each chord. Finally, the third measurement involves comparing the anchoring spaces related to each chord.

(a) octave (root) level:	0												(0)
(b) fifths level:	0							7					(0)
(c) triadic level:	0				4			7					(0)
(d) diatonic level:	0		2		4	5		7		9		11	(0)
(e) chromatic level:	0	1	2	3	4	5	6	7	8	9	10	11	(0)

Figure 5. Anchoring Space (Diatonic Space) (Lerdahl and Krumhansl 2007)

Borrowing from Bharucha's (1996) melodic anchoring space, as shown in Figure 5, the anchoring space represents relative strength of a chord and its implied diatonic space. In this third calculation, Lerdahl sums the total difference between each of the chord's corresponding spaces.

As opposed to the chord function and distance theories, research has also demonstrated the exclusive importance of musical schema, in this case longterm chord transition probability. Investigating the concept of learning through passive music listening, Jonaitis and Saffran (2009) designed a series of three experiments involving original artificial musical grammars. The first two experiments dealt with short-term learning, occurring after listening to only 4-10 phrases in one session. In these experiments they found subjects could easily and repeatably identify transitions. The third, long-term, scenario had subjects listen to 100 phrases over the course of two days. Unlike the previous experiments, in the long-term case they found subjects could identify errors within the original artificial language. These results support the theory that even passive listening allows the brain to extract and learn rules governing a system by gather statistical information about event probability.

Building on the concept of probability shaped musical expectations, François Pachet (1999) modeled harmonic expectancy through the development of a statistic based generative algorithm. Focusing on jazz music, for its rich harmonic context, he developed a system that could extract chord transition heuristics based on automatically gathered statistics. More importantly, he wanted to develop a system that, while generating syntactically correct music, could produce surprising and coherent progressions. Testing the design, he input a source of examples from a set of Charlie Parker transcriptions and others from the Real Book into his algorithm. He then compared and showed similarities between the generated heuristics and those commonly accepted in jazz theory. Additionally, he demonstrated how certain generated substitution rules could produce these surprising and coherent progressions.

2.1.3 Rhythmic Stability

Aside from choosing which pitches to generate, my tension model must also determine when to play each note. It does this through a model of rhythmic stability. Perception studies have shown

the importance of the intervals between note onsets on the stability of music (Deutsch, 1986; Desain, 1992). A seminal work in rhythm perception suggested that we perceive rhythm categorically, as a ratio of durations (Povel, 1981). Testing this theory, Povel conducted a rhythmic production study. He presented subjects with a repeating pattern of two durations. The ratio between the durations varied from .25 to .8. For consistency the sum of the two durations always summed to 1000 ms, for instance the .8 ratio consisted of 444 and 556 ms durations. After hearing the pattern, the stimuli ended and participants continued by tapping the pattern.

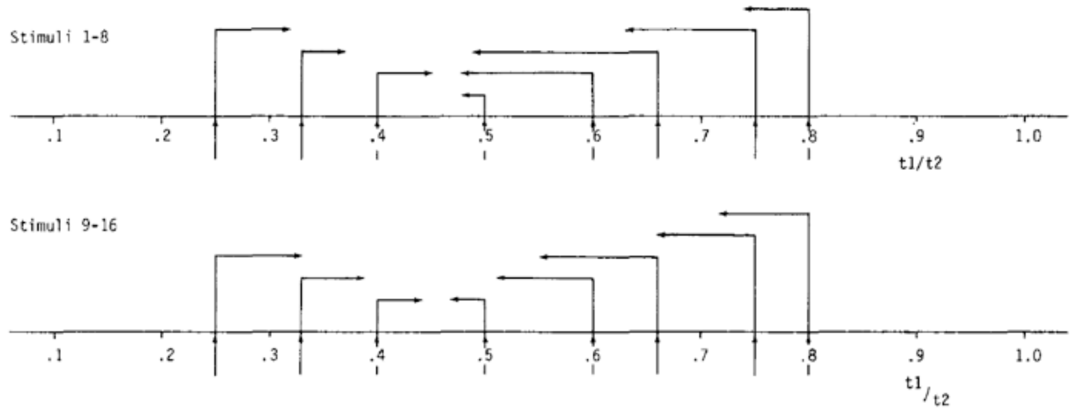

Figure 6. Divergence from the stimuli duration ratio (x-axis) and the mean produced ratio (the endpoint of the arrows) (Povel, 1981)

Table 2. Results showing mean, standard deviation, and drift in participant responses (Povel, 1981)

Stimuli				Imitations			Stimuli				Imitations		
	$t_1 + t_2 =$ 1,000			t_1/t_2				$t_1 = 250$			t_1/t_2		
No.	t_1	t_2	t_1/t_2	*M*	*SD*	Drift	No.	t_1	t_2	t_1/t_2	*M*	*SD*	Drift
1	200	800	.25	.33	.06	−.03	9	250	1,000	.25	.33	.06	−.02
2	250	750	.33	.37	.07	−.03	10	250	750	.33	.39	.07	−.03
3	286	714	.40	.45	.04	−.03	11	250	625	.40	.44	.05	−.02
4	334	666	.50	.48	.04	.00	12	250	500	.50	.47	.05	.00
5	375	625	.60	.48	.04	.01	13	250	417	.60	.51	.04	.04
6	400	600	.66	.49	.04	.01	14	250	378	.66	.55	.04	.03
7	429	571	.75	.63[a]	.19	.04[a]	15	250	333	.75	.66[a]	.18	.06[a]
8	444	556	.80	.74[a]	.19	.04[a]	16	250	312	.80	.72[a]	.17	.07[a]

Note. N = 25. Durations are in msec.
[a] These values are unreliable, as they are the result of contrary tendencies. This is reflected in high *SD*.

The results from the study show a convergence, in the imitated patterns, towards a ratio of 1:2. As shown in Figure 6 and Table 2, subjects imitated ratios smaller than 1:2 as larger than the stimuli, and represented larger ratios as smaller than the stimuli (in both cases, closer to 1:2). He interprets the data as our having a rhythmic perception strongly based on two relative categories: short and long, and these categories relate in a 1:2 relationship.

Deutsch (1986) showed that the convergence phenomenon, as demonstrated in Povel's research, results from a distortion in memory. This distortion, she argues, results as the memory stores temporal patterns as simple metrical descriptions. Supporting this theory, Deutsch conducting a comparison study. Subjects listened to a series of blips. The duration between the first two blips indicated the standard duration. A series of 'interpolation' blips proceeded, followed by a final comparison duration. Deutsch asked participants to judge the equivalency of the standard duration and the comparison duration; participants could classify the comparison duration as longer, shorter, or equal to the standard duration.

Table 3. Results from the case where stimuli comparison, C, and standard, S, are equal, and the interpolated durations are larger, smaller, or equal to the standard duration. (Deutsch, 1986)

Condition	Subcondition	Judgment	Percentage of Responses
1	C = S (I = S)	"Longer"	0.0
		"Shorter"	4.6
	C = S (I > S)	"Longer"	1.9
		"Shorter"	31.5
	C = S (I < S)	"Longer"	55.6
		"Shorter"	2.8
3	C = S (I = ½S)	"Longer"	20.3
		"Shorter"	6.5
	C = S (I > ½S)	"Longer"	19.4
		"Shorter"	37.0
	C = S (< ½S)	"Longer"	45.4
		"Shorter"	6.5

To test the effect of memory distortion, Deutsche varied the durations of the interpolated blips, I, relative to the standard, S. The study presented three major conditions: I > S, I < S, and I = S. Results from the condition in which the comparison and standard duration were equal strongly support Deutsch's theory. As displayed in Table 3, in the condition where I = S the error, incorrect "Longer" and "Shorter" judgments, was marginal. In instances where I > S, incorrect "Shorter" judgments raised to 31.5%, a highly significant result [$F(1,17) = 73.91$, $p < .001$], while incorrect "Longer" judgments increased insignificantly to 1.9%. Likewise, in the I < S case, incorrect "Longer" judgments rose to 55.6 % [F(l,17) -- 30.23, $p < .001$], while the incorrect "Shorter" judgments dropped to 2.8%. Deutsch argues the respective error rates, influenced directly by altering the interpolating durations, clearly supporting the hypothesis of memory distortion.

Building on the suppositions proposed by researchers like Povel and Deutsch, Peter Desain (1992) based his decomposable theory of rhythm perception, DECO, on the assumption that we perceive the stability of notes based on a ratio of inter-onset-intervals.

He defines the term 'binding' as the relationship between the middle of three onsets (an onset binding previous and following inner onset intervals). He hypothesized that simple ratios between bound inner onset intervals (IOIs) created greater stability than more complex ratios. Furthermore, he proposed that for near-integer IOI ratios, the stability could be modeled by a Gaussian curve centered on the exact integer ratio

In “Rhythmic Stability as an Explanation of Category Size” Desain and Honing (2002) extend DECO by further hypothesizing that the stability of a pattern may correspond to the geometric mean of the stability of the onsets. They also conducted a study to evaluate DECO and this addendum hypothesis. In the study, participants listened to 210 different rhythmic response patterns each consisting of four onsets. They asked participants to categorize each of the rhythmic patterns using a music notation-based interface. In theory, they suggest, trends should occur in which the ease of identification of patterns increases as the patterns become more stable, being made up of simpler integer ratios.

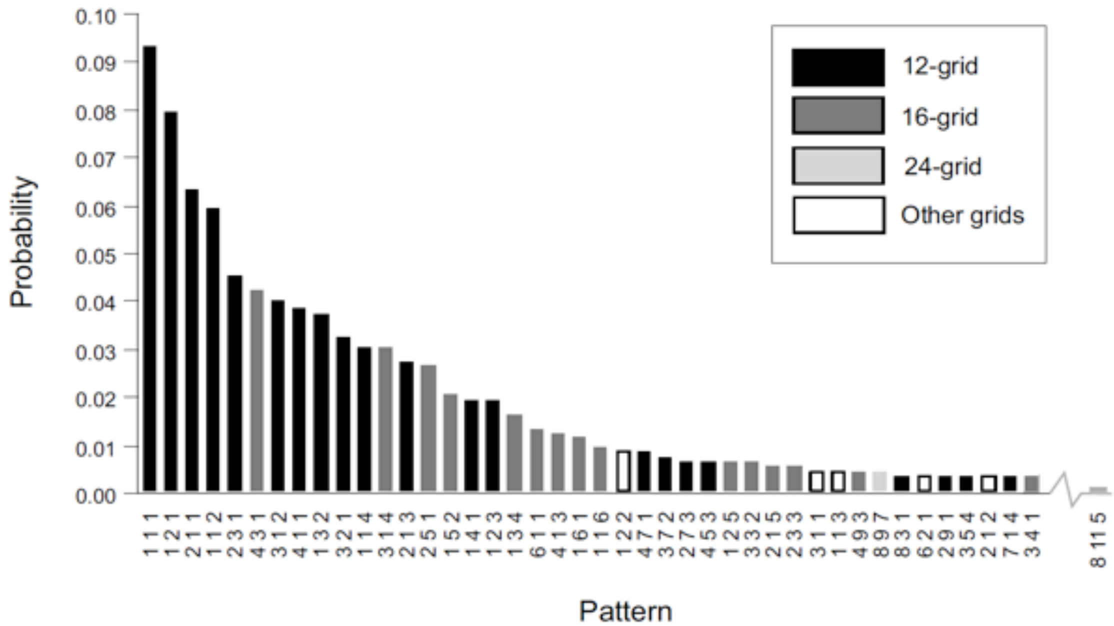

Figure 7. Distribution based on the proportion of all categories chosen (Desain and Honing, 2002)

The results from the study supported the DECO theory. As indicated in Figure 7, subjects more often tended to classify rhythmic patterns as simpler patterns. For instance, people most often classified the patterns as all three IOIs equal in duration, [1 1 1], the most stable pattern according to DECO.

2.2 Generative Music

Originating in the early 1950s, computer-based generative music branched in several different directions. The probabilistic generative approach we take in this project can be related to the pioneering work of Hiller and Isaacson, who premiered their first major algorithmic composition "Illiac Suite for String Quartet" in 1957. Written entirely by generative code, their algorithm used Markov chains to probabilistically generate note values (Belzer et al. 1981). After generating a note in each of the iterations, the algorithm tested it against a set of compositional rules. If the note passed the test, the algorithm accepted it and began generating the next note. If the proposed note failed the test, the algorithm erased it and generated a new note that was tested by the rules again. Ultimately, the work produced melodic and even contrapuntal examples that followed certain voice leading principals yet arguably lacked higher levels of musicality. This seminal work did, however, pioneer the way for computer-generative works that would follow years later.

The proceeding sections divide related work in the field of generative music into contextual and influential works. The contextual works section provides a historical background for works that directly influenced my design. This provides a framework for the succeeding section, which describes works that directly influenced the design of my generative system.

2.2.1 Contextual Work

Generative music includes a wide scope of approaches, not all of which inherently rely on computer-aid. Process music, a subsidiary of generative music, is often created entirely without a computer. It focuses on the revealing of a particular musical heuristic. In most cases some simple music material unravels over time, the unraveling being determined entirely by a single rule/operation. For instance, Brian Eno's Music for Airports relies solely on tape phasing to generatively create a continually changing texture (Tamm, 1995). In the piece, tracks of different lengths are looped continually; the different lengths are cut so that over the course of the piece they never resynchronize to their original orientation, creating the illusion of an ever-evolving texture. Attaching to this single operation, however, limits the dynamic behavior of Process Music, behavior necessary in my case to map to the realtime fish movements.

In the opposite extreme, the generative work of Iannis Xenakis focuses intently on the dynamic mapping of musical parameters to stochastic methods. In his work ST/10-1

080262, Xenakis employs his ST program, allowing him to apply probability distributions (normal distributions in this case) to various parameters (Keller and Ferneyhough, 2004). With his compositional model of when, where, and how to use different distributions for each musical parameter, ST generates a musical score. The score is then handed to human musicians for performance. While the dynamic behavior of this approach could clearly translate to the dynamic behavior needed to facilitate sonification, it's reliance on a general model make it better suited as an approach for composition.

Zicarelli's M and Jam Factory blend the space between composition and dynamic texture creation (Zicarelli, 1988). Rather than needing a compositional model his work employs realtime user input. The process begins with the user creating a sequence of notes. His systems take this sequence and pass it through a series of modules. Each can perform various transformations on the sequence of notes, including reordering them altogether. From the system, music is continuously created based on a combination of this input sequence and the operations performed by the various modules. Throughout the continued generative process the user can manipulate parameters of the various modules in order to shape the resulting music. This demands that the user make the decisions controlling the high-level features of the composition, an aspect that must be managed autonomously in my system.

2.2.2 Influential Work

The following works, unlike the previous, directly influenced the design of my system. These generative systems, like my own, give simultaneous attention to both high and low level aspects of the music they create. David Cope's "Experiments in Musical Intelligence", sought to capture both high- and low-level features of compositions in order to generate stylistically authentic reinventions of music (1991). His early work in this field in the 1980s, revolved around the concept of defining a set of heuristics for particular genres of music and developing algorithms to produce music that recreates these styles. By Cope’s own account, these early experiments resulted in a 'vanilla' music that technically followed predetermined rules yet lacked ‘musical energy’ . His succeeding work built on this research with two new premises: every composition had a unique set of rules and the algorithm determined this set of rules autonomously; this was in contrast to his previous implementation where a human realized the rule-set. The work ultimately relies on pattern recognition for analysis and recombinancy for synthesis, in an effort to create new musical material from pre-existing compositions. However, while this implementation produces effective reconstructions true to the form of the original composition, it does not have the ability to generate music in real-time.

Belinda Thorn and François Pachet each uniquely developed software that addressed the challenges of real-time generative algorithms with authentic musicality. Thorn (2001) completed the first generation of Band-Out-of-the-Box (BoB). Her work relies on two models for improvisational learning. First, with previous knowledge of the work's harmonic structure, an offline algorithm listens to solo improvisations and archives probabilistic information into histograms. Then, in real-time, BoB analyzes a human player’s solo improvisation for modal content. Based on this content and the offline-learned information, BoB then generates its own solo improvisation. From here, in the classic jazz tradition, both human and computer trade fours (each taking turns individually improvising for four bars of music) for the remainder of the performance. While providing real-time improvisation and doing so in a nearly human manner, previously determined harmonic structure limits the work's versatility. Pachet's Continuator on the other hand, builds on human performance harmonic and melodic content to generate improvisational responses (Pachet, 2002). The Continuator employs a series of Markov chains to uniquely define voice leading used throughout a segment of human improvisation. These chains, combined with the detection of the improvisation's chord content that is based on discrete time segmentation of note clusters, allow the algorithm to seamlessly continue and build upon human performance.

Similar to all of these works, our algorithm uses weighted probabilities to generate music. We have continued in the tradition of Pachet and Thorn in developing a real-time generative algorithm. However, unlike Thorn's BoB and Pachet's Continuator, which rely heavily on live human performance to drive their real-time generation, our autonomous process uses parameters determined by dynamic visual information, the movement of fish, as input. In addition, our work develops unexplored areas of generative musical tension as described below.

2.3 Sonification

Sonification systems translate non-audible data into the world of sound. The history of sonification generally demonstrates a split in its focus. Most often creators of sonification systems, intentionally or unintentionally, choose between building systems that focus on either representing non-audible data in an audible way or solely for driving some musical aesthetic. In recent history, more and more works have begun blending the two foci. The three following sections will distinctly outline each of these three approaches.

2.3.1 For Data Representation

The use of sonification as a tool for complex data display continues to become increasingly popular. Such displays can represent complex data sets both of significant depth and at high rates, in a manner that cannot be captured by visual display. Sonification applications can be as simple as Geiger-counters clicking to reveal radiation levels. A more complex example is the sonification application used by NASA's Voyager II mission that represented visually unclear data by passing it through simple sine wave audio generators. As a result, this basic sonification of the data revealed micrometeoroids (Kramer et al., 1997). In the field of medicine, sonification of image based MRI and EEG scans can help facilitate detection and identification of unhealthy parts of the brain (Hermann 2002). Sonification, in this case, enhanced visual data by revealing subtleties often missed by visual cues.

While sonification maintains an important role as a tool in technical fields, it also serves as a valuable aid to the visually impaired. Sonification for the visually impaired often finds its place where haptic feedback fails in providing easy and clear translation of visual information. Projects such as SoundGraphs, AudioGraph, and the Sonification Sandbox describe the data from graphic diagrams through means of continuous MIDI pitches changing over time (Kennel 1996; Brewster, Ramloll, and Yu 2000; Mansur 1985; Davison and Walker 2007). In their Mathtalk project, Edwards and Stevens brought sonified mathematics to the visually impaired through the use of synthesized speech and voice recognition (Edwards and Stevens 1997). Similarly, with a combination of synthesized speech and simple controlled pitch synthesis, several projects have taken on the task of describing visual geographic data, through use of GPS devices, interactively guiding people through cities (Boll et al. 2006; Wilson, Walker, Lindsay, Cambias, and Dellaert 2007). While these projects clearly relay visual information with audio cues, most employ a utilitarian approach to sound design that doesn't attempt to represent high-order abstract information.

2.3.2 For Musical Composition

Many sonification-based musical works use external data as input to automated composition algorithms. Luke Dubois' Hard Data, for example, employs various statistical data from recent U.S. military actions in Iraq, everywhere from civilian deaths to fiscal year budgets, to drive its compositional process (Dubois 2009). Hard Data does not inform nor yield extended sensing with this data, but rather uses it as an input for the composition.

Like Hard Data, Rubin and Hansen's Listening Post (Rubin and Hansen, accessed 2011) focuses less on accurately representing non-audible data and more on raising general awareness about an issue. In their case, the installation work Listening Post sonifies the traffic through Internet chatrooms and bulletin boards. They sonify text posted on these public sites as libretto sung by voice synthesizer. In doing so they hope to raise awareness to the "content, magnitude, and immediacy of virtual communication". While having a very tight mapping, literally singing the text that is printed, the work does little to musically represent the data in a way that it cannot be represented otherwise. Similarly, Dunn and Clark sonify protein data in Life Music. In this work, they use amino acid sequences found in proteins to inspire the compositional process (Dunn and Clark, 1999). They attach the solubility of the various amino acids to relative pitch and preserve their sequence as contiguous chains of notes. Ultimately the data, amino acid sequences, only loosely affects the composition by affecting relative pitch in the melodic sequences, leaving all other creative compositional choices to the human artist.

2.3.3 For Data Representation and Musical Composition

Sonification through music can provide not only an aesthetic medium that can be entertaining and engaging for the listener, but also clearly represent non-musical data (Dodge, 1962; Flowers et al., 2001; Polli, 2004), especially those that cannot be represented in a graphical manner. An early yet seminal example is Charles Dodge's Earth's Magnetic Field. In the work, Dodge musically maps changes in Kp (magnetic field) readings from magnetic observatories (Dodge, 1962). Typically observatories record these readings graphically, with levels represented vertically and time horizontally. In a more representative interpretation, Dodge maps the level of the Kp reading to relative pitch and timing of the events to temporal organization of the piece. Furthermore, he more inclusively ties tempo, dynamics, and register to "both the larger and smaller dimensions".

Flowers et al. (2001) conducted research in sonifying historical weather data. They point out that the weather data represented as a “multivariate time series” is ideal for musical sonification, which could be described the same way. The work intends to represent the data in a way that enables comparison of historical weather data across different years in effort to reveal trends. In order to simultaneously represent different categories such as temperature, rainfall, or snowfall they mapped each category to unique instrumentation. Mappings such as precipitation amounts to note density are clearly representative while also naturally shaping the aesthetic of the work. Time scales also mapped from the historical weather data to the music, with precipitation durations represented by note durations and the month-to-month time span equally mapped structurally across the composition.

In Atmospherics/Weather Works, Polli also sonified meteorological events, storms in particular (Polli, 2004). Unlike the work by Flowers et al., Polli's sonification focuses on more short term, rather than historical, changes in weather and tries to additionally capture visceral reaction to these changes in her music. She focuses not only on the aesthetic application of the information but also emphasizes potential use of it to reveal aspects of the data less evident in graphic visualizations. In describing the potential utility of music-based informative sonification Polli explains: "Through an effective sonification, data interpreted as sound can communicate emotional content or feeling, and I believe an emotional connection with data could serve as a memory aid and increase the human understanding of the forces at work behind the data." Polli suggests not only that sonification can carry continuous high level information, but that it can reveal more about a source than the sum of the data. Similarly, in our research I hope to not just translate visual information as music but to do so in a way that reveals how we perceive that information. I attempt to merge the ideas of sonification and generative composition, addressing both the utilitarian and aesthetic goals of sonification by giving attention to multiple low- and high-level mappings of the non-musical to musical spaces.

CHAPTER 3

ACCESSIBLE AQUARIUM

The most recent implementation of the Accessible Aquarium project employs the generative music system detailed in this thesis. The project aims to make "informal learning environments", like an aquarium, accessible to the visually impaired and to improve the experience of such exhibits to the sighted. To do so a computer vision team develops custom software to track the fish (Pendse et al., 2008). This tracking data informs a sonification system that maps it to music. Early attempts included several models of either simple or more complex mappings. Simple models mapped low-level features such as absolute position and pitch height. More complex models mapped voices from precomposed works, like Bach chorals, to individual fish; the voices were filtered and their amplitudes enveloped to relay gestural information about the fish movements. Informal evaluation of these early attempts of autonomous sonification systems provided mixed results. Users rated most as aesthetically "acceptable". However, a clear divide classified the simple mappings as relating to the source material but being musically uninteresting while the more complex were musically more interesting but often confusing in the mapping to the fish (Walker et al., 2007). Before further extensively developing sonification and tracking systems, however, research in the project focused on identifying salient fish movements in the aquarium (or in any dynamic exhibit in general). This research evolved into a formal subject-response study.

The study included 13 sighted university students. Participants watched a projection of three videos: a real aquarium, a simulated one, and an ant video. A "Think Aloud" portion of the study played the videos silently and asked participants to describe what they saw continuously as the video ran. Data collected consisted of quantifying the mean number of times participants mentioned each feature.

Table 4. Results and analysis of mean mentioning of features (Pendse et al., 2008)

Creature	Grand	Std	Real	Sim	Real	F	p	
Location	4.05	0.50	5.46	3.92	2.77	6.951	0.006	*
Color	3.69	0.37	2.00	8.62	0.46	56.671	0.001	*
Size	2.74	0.38	5.92	1.85	0.46	25.557	0.001	*
Species	2.28	0.55	2.85	3.92	0.08	5.679	0.017	*
Enter/exit	2.18	0.38	2.46	1.08	3.00	4.361	0.029	*
Direction	1.90	0.38	1.92	3.00	0.77	6.584	0.014	*
Behavior	1.85	0.36	0.69	2.46	2.39	3.120	0.080	~
Background	1.59	0.31	2.00	2.00	0.77	3.361	0.053	~
Interacting	1.10	0.12	0.69	0.92	1.69	3.852	0.055	~
Grouping	0.87	0.13	2.15	0.08	0.39	17.818	0.001	*
Speed	0.69	0.20	0.85	1.00	0.23	2.383	0.121	
Liveliness	0.44	0.22	0.23	1.00	0.08	2.362	0.145	
Surround	0.39	0.15	0.77	0.23	0.15	3.369	0.075	~
Shape	0.26	0.11	0.54	0.15	0.08	4.326	0.051	~
Feeding	0.18	0.06	0.23	0.15	0.15	0.103	0.896	
Sound	0.08	0.06	0.00	0.15	0.08	1.565	0.233	
Acceleration	0.00	0.00	0.00	0.00	0.00	-	-	

They created the Grand category list out of features mentioned repeatedly by subjects. It lists the mean number of mentions of a feature, across participants. Some of these features varied across videos. The results of a repeated-measures ANOVA showed several significant results amongst the mentioned features across videos, indicated with asterisks. The important results of the study showed that location, color, size, species, enter/exit, direction, behavior, background, and interaction were most important to viewers (Pendse et al., 2008).

I address the significance of these features and the shortcomings of the previous sonification implementations in the mapping of the generative tonal tension Model to the Accessible Aquarium tracking of fish. My mappings emphasize the low-level visual features noted as significant in the previous study while also addressing more complex fish behavior with high-level musical features.In my system, individual voices map to each fish in the aquarium. This helps to provides the listener with a detailed musical description of the aquarium, rather than a vague sense of the overall activity. Instrumentation of each voice parallels the species, and indirectly the color and size which tend to correspond to species. For example, the clown fish, which are all orange and white in color and medium sized, are all represented by marimba-like sounds. I found this mapping intuitively captures the entire scene. Given the importance of position, found in the previous Accessible Aquarium psychological research, my system represents absolute position of the fish in simple one-to-one mappings. The system maps the two-dimensional location (limitation of the computer-vision tracking) of the fish to a two-dimensional musical space. Even with the limitation of a stereo setup, the x-axis position of each fish is intuitively represented by the panning position of the music generated to represent that fish. Without the aid of vertical panning, I chose to represent the y-axis position of the fish with the relative pitch height of the notes generated. Through personal observation, I have found that the combination of x- and y-axis mappings provide a relatively clear sense of absolute position of a fish in the aquarium. The speed of the fish maps to a subdivision of the tempo in the melody, so that denser melodies correspond to faster fish movements. This lower level mapping reveals higher level information about the fish movements. For instance, as a fish accelerates or decelerates the notes gradually become more or less dense, respectively. A sudden burst in speed corresponds to a burst of dense notes.   As described earlier, research also showed the importance of the behavior of the fish. Trying to analyze this from the positional tracking, I measured the rate of changes in direction. I found that this rate indicated how erratic the fish moved, as erratic movements generally involved rapid changes in direction. I mapped this higher-level descriptor of the fish movements to the musical tension, with higher erratic behavior mapping to more tense musical results.

CHAPTER 4

HYPOTHESIS

I propose that combining modified versions of several analytical models of musical tension can create a comprehensive generative model of musical tension. Furthermore, I propose that this generative model of musical tension can effectively be applied to the task of sonfication, addressing both high- and low-level mappings of non-musical to musical information while simultaneously attending to musicality. I propose two criteria for assessment of the generative system outlined in the thesis.

1. Tonal tension modeling: The system will address music generation with respect to the creation and resolution of tension. To do so, it will address three key features shown highly correlated to the perception of tension in music: melodic attraction, harmonic expectation, and rhythmic stability (Farbood, 2001). I will evaluate the effective generation of tension in a subject-response study. In the study, subjects will rate the relative perceived tension of musical excerpts generated by the system. I will use a cross-correlation analysis to assess the effectiveness of the algorithm, comparing the predictive input tension generated by the algorithm and perceived rating provided by subjects. In the evaluation I will also examine perceived tension across pitch range (excerpts with pitches above or below a threshold) and across various instrumentation, both features also having been shown to highly affect tension, but were not included in the model because of the lack of quantified research regarding their effect (Helmholz, 1885; Plomp, 1964; Hutchinson and Knopoff 1978; Farbood, 2001; Vassilakis, 2007). This evaluation will allow us to normalize the system by accounting for these features.

2. Sonification: In applying the generative model to the sonification of the Accessible Aquarium Project, I will address the dichotomy between representing data sonically and creating a musical product. Approaching this divide, I propose a simultaneous mapping

of both high- and low-level features. I conducted an online survey to evaluate of the sonification task. The survey asked subjects to assess video excerpts of several different implementations of the Accessible Aquarium Project, including the latest with the algorithm described in this paper. The assessment addresses aspects of musicality and information representation of each sonification system. In addition to previous implementations of the Accessible Aquarium project, the evaluation compares responses to separate excerpts of my sonification system in which the audio and video are aligned and misaligned (as a baseline, disrupting the original sonification while maintaining the same general musical sounds and aquarium visuals)

CHAPTER 5

METHODS

The primary goal in the design aimed to provide a stable groundwork of relatively independent modules, capable of continued additions and evolution. Appropriating for this, our design focused on the development of a simple yet robust algorithm. It is simple in the sense that the design does not rely on a complex network of rules and conditions; it is robust in that the music produced by the algorithm should be capable of affectively representing a wide array of musical gestures. In order to accommodate the need for our music system to operate in realtime, I designed and implemented the generative components in Max MSP.

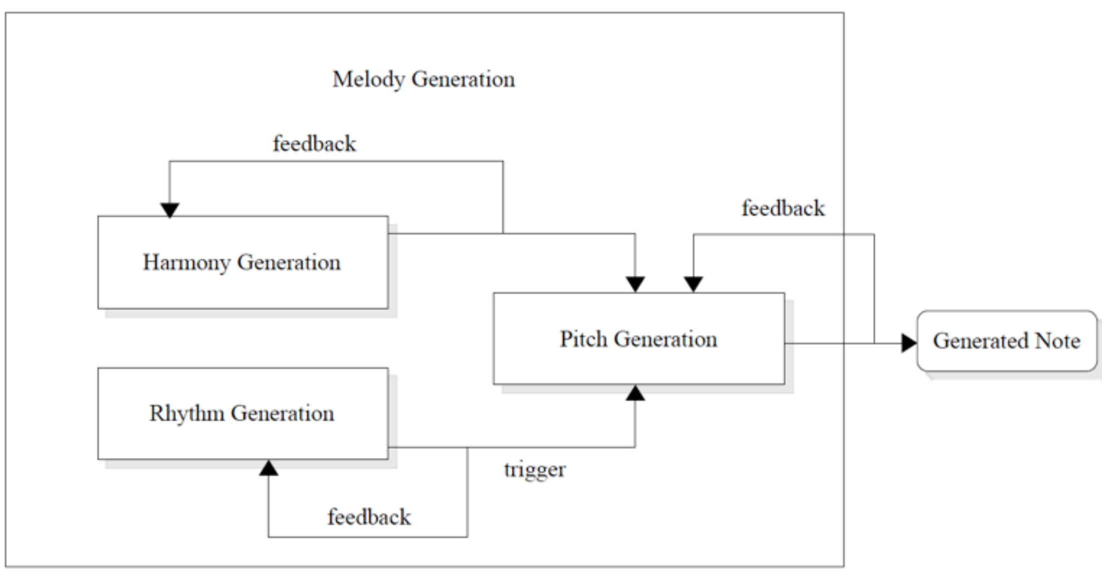

Figure 8. The interaction between rhythm-, pitch-, and harmony- generation modules in order to generate the 'next' note

The design consists of three major components: pitch-, rhythm-, and harmony-generation, as shown in Figure 8. The three modules each interact to ultimately generate the notes of a single voice melody. Rhythm generation determines the onset of notes. It's output triggers pitch generation, which determines the pitch based on the current state of the Harmony generation. Output from Harmony informs Pitch by generating chords consisting of anchoring tones to which pitches are attracted; this will be explained in further detail later. All three modules behave as state machines relying on feedback of the previous state to determine the next.

In the context of applying the generative algorithm for sonification, I drive these generative modules with input from computer-vision tracked fish. With images from a single Prosillica camera, our OpenCV-based system works with models of fish based on the general size, shape, and color of each. This allows it to effectively identify and track their independent movements. Using this data, I have focused the design of our generative music algorithm to represent the experience of viewing the aquarium.

In order to address mapping both low- and high-level visual parameters to musical parameters, I segmented the various attributes of the visual information I wanted to represent. On the lowest level, I decided to convey simple location-based information such as the position of a fish at any given point in time. Additionally, I wanted the sonification to depict gestural information about their movements by mapping the speed of their gestures to the rhythms of the generated music. With respect to erratic versus non-erratic behavior , I decided to represent the general ambiance in the aquarium by changes in harmonic expectancy and individual behavior such as predictable or erratic swimming patterns of the fish by relative melodic tension. The latter led to the design and application of the generative tension algorithm described in this paper.

CHAPTER 6

IMPLEMENTATION

Table 5. Sonification mappings table.

Mappings	
Fish	Music
Y-position	relative pitch height
X-position	panning
speed	relative note density
erratic behavior	tension (melodic, rhythmic and harmonic)
timbre	species

6.1 Tracking Visual Tension

One of the goals of the application of the algorithm in sonification included mapping visual tension to musical tension. In order to detect visual tension I developed a measurement to describe the flow of tracked fish movement. This calculation maps lower numbers to consistent movements and higher numbers to erratic movements (It is important to note that by describing "erratic" movements. I am not making claims about the fish psychology, only about the unpredictability of the movements themselves). I define erratic and unexpected movement by the density of changes in direction. Research (Berthoz, 2000; Hagendoorn, 2000; Engel et al., 2001) supports this claim, as Hagendoorn writes "as deviation from and correspondence between the actual motion trajectory of a moving object and the trajectory as predicted by the brain of the observer, gives rise to two distinct emotional responses, analogous to the euphoria and frustration of catching or missing a ball" (Hagendoorn, 2004, p. 80). Since we naturally and unconsciously make predictions about the fish's movements and these predictions are violated by these rapid changes in direction, I conclude that by measuring the density of these rapid changes we get a sense of the visual tension created by the swimming patterns.

The first difference of each coordinate indicates a direction vector. Comparing this direction vector to the previous one reveals whether or not the tracked fish has changed direction. The summation of the number of changes in the tracked fish's direction over a given period time provides the expectancy of its movements. In particular I use a running sum over a period of three seconds. A maximum threshold of ten changes in direction, across the running sum, maps to a maximum visual tension level and zero changes in direction maps to a minimum visual tension level (Note, since the computer-vision tracking operates at approximately thirty frames per second it does not conflict with this metric). I did not include the magnitude of the direction changes as, by my own intuition, this simple running average captures erratic behavior of the fish. These visual tension values map directly and linearly to the input tension parameter of the generative music tension algorithm.

In order to represent this measurement sonically, I use it to control the tension levels, which influence the generation of harmonic, melodic, and rhythmic features. Thus, as the tracked fish changes from flowing movements to disjunctive movements, the music changes from less to more tense.

6.2 Rhythm Generation

I based the Rhythm Generation module on the model proposed by Desain and Honing, for analysis of rhythmic stability (Desain and Honing 2002). Their work demonstrated the relationship between rhythmic stability and the bounds between contiguous inter-onset intervals (IOIs). In particular, they showed direct proportionality between the complexity of ratios between contiguous durations and their relative rhythmic stability.

Extending the concept for analyzing stability into a predictive model, I implemented a method for rhythmic generation. In our predictive implementation, the algorithm refers to previous IOIs to inform the generation of the future onsets, as shown in figure 9. Specifically, provided a high or low input tension level the algorithm accordingly gives preference to future onsets that form either complex or simple ratios, respectively, with the previous IOI.

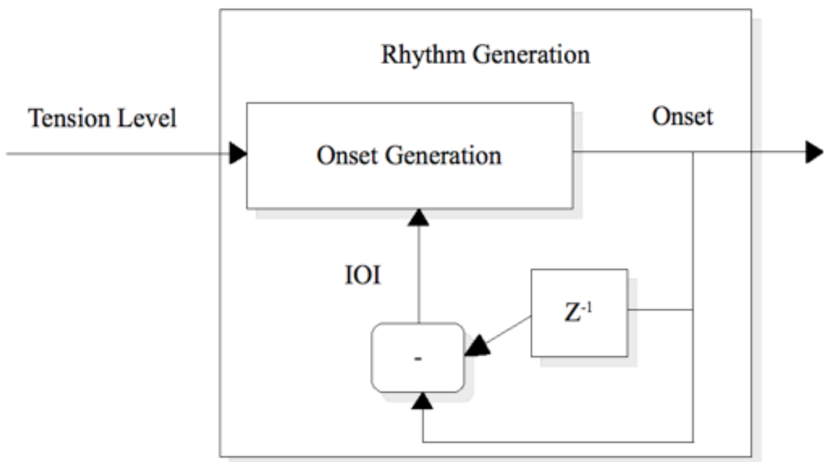

Figure 9. Rhythm Generation based on Tension Level and previous Inter-Onset Interval

The onset prediction relies on a lookup table in order to pseudo-randomly generate future onsets. Its lookup table includes a list of ratios arranged according to complexity, where ratios such as 1/2 and 2/1 that were shown in previous studies to be more stable occur low on the list while 9/2 and 2/9 occur significantly higher. Influencing the pseudo-random generation, high input tension values give weight to ratios high on the list while, vice-versa, low-tension values give weight to lower ratios. These onsets are ultimately mapped onto a constant metrical structure.

In this sonification context, I continuously map the speed of the fish movements to the note density, as shown in figure 10.

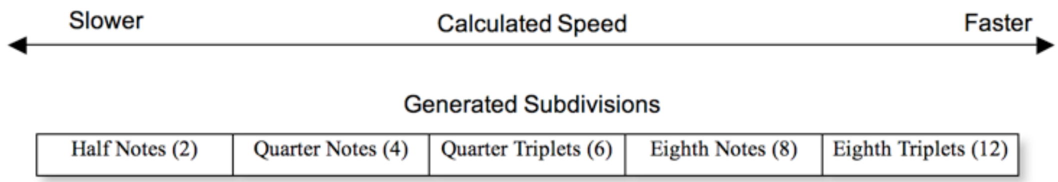

Figure 10. Mapping speed to note density

In this case, the algorithm combines the note density mapping with the rhythmic stability prediction. To do so, the algorithm first considers the influence of the speed mapping. This determines the relative note density. The onset generation then pseudo-randomly generates the next onset with a more or less complex ratio between IOIs, but also weights the lookup table probabilities based on distance from the relative note density. As such, a fish's speed maps directly to the density of notes and the visual tension maps to input tension value of rhythmic stability, as described before.

6.3 Melody Generation

I developed a method for pitch generation that could controllably change melodic stability and tension in real-time. I based our method of melody generation on Fred Lerdahl's theories of tonal pitch space (Lerdahl 2001). Compared to similar work in the same field (Margulis 2005; Narmour 1992), Lerdahl's research in cognitive theory addresses the concepts of stability and tension in detail. While Lerdahl originally intended this equation as a theoretical means of deciphering relative stability, Deliège described these formulas as unproven and bearing limited usability as an analytical tool (Deliège 1997). However, it has been shown recently that these formulas can be used effectively in a generative and interactive manner (Farbood 2006; Lerdahl and Krumhansl 2007).

My implementation is based on Lerdahl's analysis of voice leading, which depends on two major components: anchoring strength and relative note distance. The concept of anchoring strength maintains that given a certain pitch space there remain areas of greater and lesser attraction.

Table 6. Relative note distance

Gb	G	Ab	A	Bb	B	C	C#	D	D#	E	F	F#
7	6	5	4	3	2	1	2	3	4	5	6	7

For example, the pitch C in the center of the Table 6 represents the previous pitch. The relative note distance grows as notes move farther away from C. This distance inversely affects the probability of that following note; for instance, C to C has a 'distance' of 1 to avoid division by 0. Accordingly, this value controls the generative preference towards smaller melodic intervals.

In Lerdahl's stability equation for voice leading the effect of the next note's stability is inversely proportional to the previous note's anchoring strength. Building on the criticism of the subjective nature of Lerdahl’s formulas, we decided to experiment and manipulate some of the parameters in the formula, in an effort to reach our own subjective satisfactory musical results.

$$L = \left(\frac{a_2}{a_1}\right)^z \left(\frac{1}{n^y}\right) + x$$

$T = 0 : 1$ $a_x = 15 : 1$
$z = 2 : 0$ $n = 0 : 12$
$y = 1 : 0.1$
$x = 100 : 10$

Equation 2. Altered form of Equation 1. for generative purposes; L represents the likelihood that a given pitch will occur next, T represents the input tension parameter which scales linearly to z, y, and x to affect the likelihood of stable versus unstable pitches

As shown in Equation 2, we added variables x, y, and z. The input tension, T, maps to these variables, controlling whether stable or unstable pitches are more likely to be generated. The larger the variable is, the more likely it is for an unstable pitch to be played. Changing z controls the influence of anchoring strength in determining the next pitch. As tension T increases, z decreases, reducing the likelihood of strong anchoring pitches to be generated. Similarly, y effects the impact of the relative step-size. As discussed earlier, theorists have shown that smaller steps between pitches increases the perception of more stability. As the tension input value approaches zero, a small pitch step size becomes more likely, and therefore the output becomes more stable. Variable x affectively adds noise to the equation. By raising x, anchoring strength and step-size become relatively less significant in generating the next note. This makes unstable pitches more likely.

We empirically derived the mapping from input tension to variables x, y, and z. Through trial, error, and tweaking all three parameters, we gradually found a range for each that intuitively corresponded to the input tension values.

6.4 Harmony Generation

The harmonic component of the system does not directly generate notes. Instead, it informs the pitch generation system. The harmonic generation system outputs harmonic changes in the form of diatonic anchoring spaces, tables giving proportional weighting to chord and scale tones . Currently harmonic changes occur regularly, once every four beats. As the notes produced by the pitch generation tend towards the anchoring tones, the music created by the system will follow the harmonic structure output from the harmonic generation.

The harmonic generation system involves the combination of two models. The first includes a generative model of the harmonic analysis theory presented by Lerdahl in Tonal Pitch Space. His evaluation of the model demonstrated its ability to effectively analyze the tension created due to harmonic changes. However, empirically, I found that my generative implementation of this model tended to produce uncommon chord changes, far more often than experienced than in a normal review of musical literature. For instance, in Lerdahl's model, the cost of a movement from a I chord to bVII is relatively small, however that chord change rarely happens compared to a I to II, which would have an equivalent cost in his system. Though the movements in these changes corresponded to similar pitch spaces, the surprise in defying expectations of a more common changes seems to create additional tension. To account for this, I designed a second statistic-based model. This model accounts for the likelihood of any chord moving to any other chord, based on a review of musical literature.

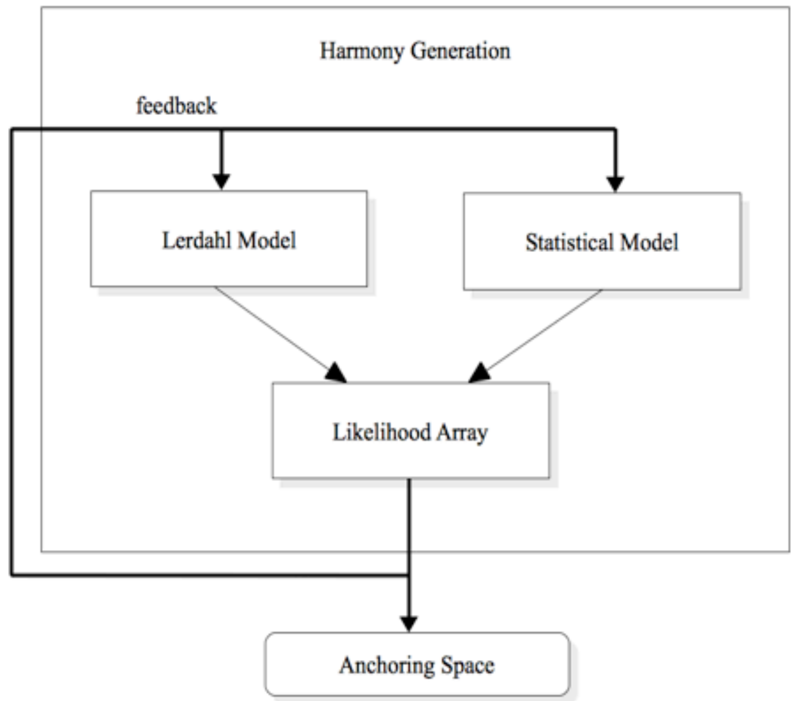

Figure 11. Harmony Generation Diagram demonstrating the combination of two models and feedback informing future generation

For both models a program was written that produces matrices describing the probability of moving from any one major, minor, or diminished triad / diatonic space to any other. The probability of each space from both models are multiplied. Using them generatively, the last space output by the harmonic generation system becomes the input in generating the next. It assesses the transitional probability of moving to any other space and generates the next based on these probabilities.

6.4.1 Model I: Lerdahl's Equation

Lerdahl describes his model as "a music-theoretic formal model of tonal pitch space that correlates with the empirical data and that unifies the treatment of pitches, chords, and keys within a single framework". His model calculates the theoretical distance between any two chords.

> **Diatonic chord distance rule**: $\delta(x \rightarrow y) = i + j + k$, where $\delta(x \rightarrow y)$ = the distance between chord x and chord y; i = the number of moves on the cycle of fifths at level (d); j = the number of moves on the cycle of fifths at levels (a-c); k = the number of non-common pitch classes in the basic space of y compared to those in the basic space of x.

Figure 12. Lerdahl's model of the theoretical distance between two chords (Lerdahl, 2001).

The model divides a chord into a series of interrelated pitch spaces, as shown in Figure 12. Ordered hierarchically, these space draw on assumptions of the importance and stability of certain relative tones. For instance, the root and fifth of a triad serve as stronger anchoring tones than the third. His distance model operates by evaluating shared pitch-classes across these different spaces plus the number of movements between the chords on the circle of fifths.

$$
\begin{array}{cccccccccccc}
 & & & & & & & \underline{7} & & & & \\
 & & \underline{2} & & & & & 7 & & & & \\
 & & \underline{2} & & & & & 7 & & & & \underline{11} \\
0 & & 2 & & 4 & 5 & & 7 & & 9 & & 11 \\
0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11
\end{array}
$$

$$\delta(\mathrm{I}/\mathbf{C} \rightarrow \mathrm{V}/\mathbf{C}) = 0 + 1 + 4 = 5$$

Figure 13. Example of calculating distance between chords (Lerdahl, 2001)

For example, as shown in Figure 13, the distance between a I and V in the same key is calculated by a cost of 1 movement on the circle of fifths and 4 non-shared pitch-classes between the two spaces. Using this calculation, my algorithm relies on a matrix created with the cost of moving between any two chords, in the same key. Because the algorithm combines the chord distance calculation (the Lerdahl model) with my statistic-based model, I normalized the rows of the matrix. So, for each chord, the likelihood of moving to the closest related chord has a value of 1 and most distant chord has a value closest to zero.

Lerdahl argues that this distance corresponds to tension created by the harmonic movement. Examining this claim, he conducted studies comparing a subject-response to tension with his analysis of tension in Bach Chorales.

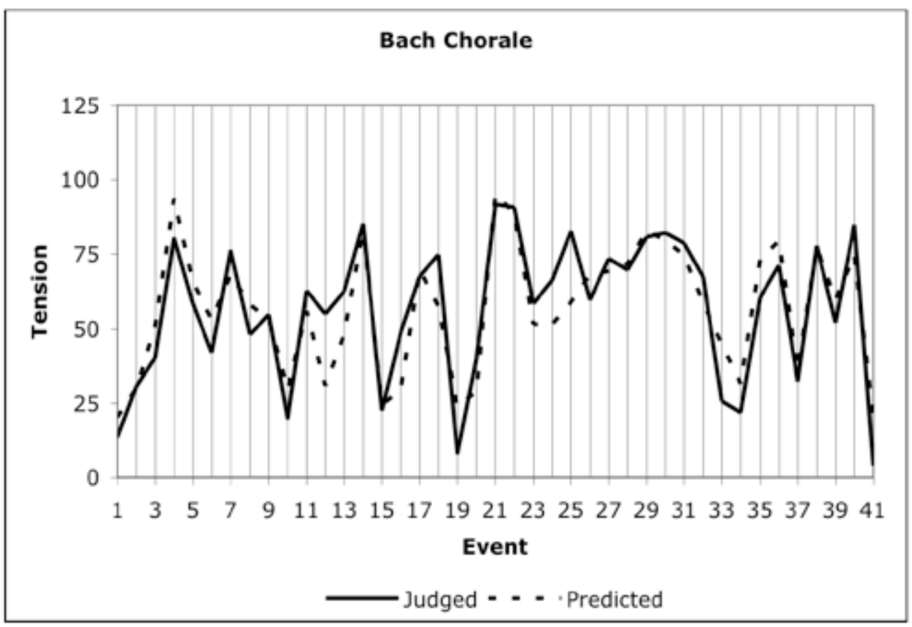

Figure 14. Results from Lerdahl's evaluation of his Tension model (Lerdahl, 2001)

In the study, he combined his harmonic distance model with his hierarchical model of tension, detailed in the Generative Theory of Tonal Music, and a model of surface tension to form a single more comprehensive model of tension. Subjects responded to tension with a single slider to indicate how tense they perceived a musical passage. Figure 14 shows the results of his study. In a multiple regression analysis he found the correlation between judged and predicted tension highly correlated with p(tension) $< .0001$ and correlation coefficient 0.95 (Lerdahl, 2001). While demonstrating the predictiveness of his entire model, Lerdahl does not evaluate individual components of the model, such as a single evaluation of the chord distance model. Furthermore, since the study only evaluates prewritten compositions, it does not comprehensively evaluate the model's predictiveness of more irregular patterns; as I suggested in the introduction, the model does not make a distinction between the novelty of a I → bVII movement compared to regularity of I → II, despite their being equally distant. To account for these regularities, I introduced a second statistically based model.

6.4.2 Model II: Statistical Comparison

The statistic-based approach predicts chord-likelihood based on a model of likelihood for any chord to move to any other . In order to do so, I symbolically notated every chord transition in the first Real Book. I notated chords by their root, relative to the key, and quality of the base triad (major, minor, or diminished). For further details refer to Appendix I. The Real Book provides a corpus of classic jazz literature. The probability of each of these transitions was then calculated based on the number of occurrences (see Appendix I). While this does not necessarily absolutely correspond to the listening schema of an average listener, of Western Tonal music, looking at the relative statistics from the literature I found that it corresponds to harmonic transition regularities. For instance, chord progressions extremely prevalent in pop music, such as I → IV → V → I and I → ii → V → I still occurred relatively more often than any other progression. Jazz music, compared to popular music, however, employs a much richer harmonic vocabulary. This vocabulary provides the model with a higher resolution between the common and uncommon changes, and provides more possibilities for chord movement in the generative implementation.

With the data of every chord change in the first Real Book, I wrote a small program to iterate through every change and build a matrix that describes the number of times any chord moved to any other chord. As with the implementation of the Lerdahl model, I normalized the matrix across rows. It assigns chord changes that occurs most often a value of one and changes that do not occur at all a value of zero.

6.4.3 Mapping to Tension

In order to map the harmony generation model to tension, the algorithm dynamically reassigns probability distributions based on a tension input parameter.

$$P(a \rightarrow b) = p(a \rightarrow b)^{(2-2\mathrm{T})}$$

Equation 3. Probability P of chord a moving to chord b given tension input parameter T and model-combined probability p.

The algorithm dynamically generates the probability of moving from the last chord to the next by raising the probability from the combined Lerdahl- and statistic-models to the tension input value, which varies from 0:1. At low value for T, the probability equation gives high likelihood to the closer related and more common chord changes. At high values, the probability of all chords approaches one, more distant and unrelated chords become closer to equally likely as close and expected chords.

6.4.4 Mapping Globally

In the Accessible Aquarium application, a single harmony generation component globally informs the melodies generated corresponding to each fish. Since the harmony generation becomes a global parameter, we input a global tension parameter into it. The global tension parameter sums the individual tension parameters corresponding to the movements of each fish, as described in section 4.1. Using this global tension parameter, the tension created by harmonic changes reflects an overall sense of visual tension in the aquarium.

Example videos with generated audio can be found here: http://ryan.nikolaidis.us/Thesis/

CHAPTER 7

EVALUATION

Two independent studies were used in order to examine the effectiveness of the system. Both of the studies relied on various subject-response evaluations. The first study assessed the effective creation and resolution of tension with regard to the combination of melodic attraction and rhythmic stability. Additionally, the study evaluated tension across several other features such as instrumentation and register; this allow me to normalize effective tension across these other features. The second study evaluates the application of the generative algorithm in the context of sonification. This study compared sonification created by my system with that of the earlier systems used in the Accessible Aquarium project. A baseline for comparison uses similar sounding music that is not generated in response to the visual stimuli with which it coincides.

7.1 Study 1: Melodic Attraction and Rhythmic Stability

In an effort to evaluate the effectiveness of the algorithm we developed in representing various degrees of tension in real-time, we conducted a user study designed to assess the relationship between algorithmically generated tension and perceived tension. The user group included one hundred volunteer students pooled from our university. To each subject we presented 100 four-second excerpts of audio. To account for the relative effects imposed by the order of the excerpts, each trial employed a randomized sequence.

7.1.1 Materials

The musical excerpts generated by the algorithm were manipulated for register, density, and instrumentation, in an effort to evaluate the influence of these parameters on perceived tension. Knowing how these other features affect the perception of tension will allow us to normalize across features in future revisions of the algorithm. Pitch material was classified as either high- or low-register, as excerpts contained notes that are exclusively either higher or lower in pitch than C4. Note density was categorized using average IOI, as either longer or shorter than 750 milliseconds. We subcategorized instrumentation by sustain and brightness levels. Two of the instruments were sine tone generated, one with long and the other with short sustain. Three other sampled instruments represented differences in sustain and brightness, classified as either bright or dark in timbre. For all combinations of these categories we generated excerpts at five different tension levels, with a tension level of 5 representing high tension, and 1 representing low tension.

7.1.2 Procedure

Subjects wore earphones and listened to a series of 100 audio clips, which were played in a completely randomized order for each subject. After listening to each clip, listeners indicated tension by typing values into a number box, in a testing program running on a computer in front of them. Subjects were asked to rate the tension of the musical excerpt using magnitude estimation (cf. Stevens 1975) . Magnitude estimation provided a solution to two major concerns. First, in an assignment system constrained by maximum and minimum values, the subject limits the range with a first assignment of either boundary. For instance, if the maximum permitted value was 10 and the subject indicated 10 for the previous excerpt yet found the next even more tense, they would have no additional range for expressing this relativity. In order to resolve this problem, the procedure could have first provided maximum and minimum examples of tension. However this would impose designer-interpreted conditions on the subjects. Magnitude estimation, and in particular "modulus-free" magnitude estimation, on the other hand, is used to address these issues. In order to account for earlier inconsistencies due to initial ambiguity in the perceived range and resolution the first five values of each trial were discarded.

7.1.3 Results

Working with data from magnitude estimation that has no consistency in range and boundary across subjects, we used geometric means, rather than normal arithmetic means, in order to represent all of the available data within an equivalent context across categories and between subjects. Precedence for this approach, using geometric means when analyzing magnitude estimation from large groups, has been established by previous researchers (Walker, 2000; Stevens, 1975). While changes in note density compared to perceived tension showed only slight correlation, registration and instrumentation proved significantly influential towards affecting perceived tension.

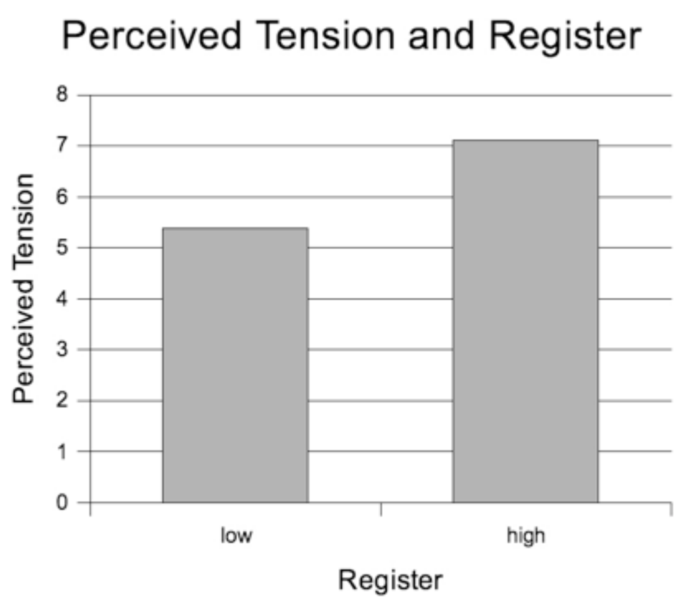

Figure 15. Geometric mean response in perceived tension as compared to change in register.

In general, stimuli generation involved creating two clips for every combination of features (including high and low register, five different instruments, and five input tension settings). Two clips were used so that every unique combination of features was played twice for the listener (without literally repeating material). As shown in Figure 15, music generated by the same parameters but in a higher register proved, on average, twenty-four percent more tense than when compared to music in a lower register. These results are congruent with Farbood's findings in her subject-response study (Farbood, 2001)

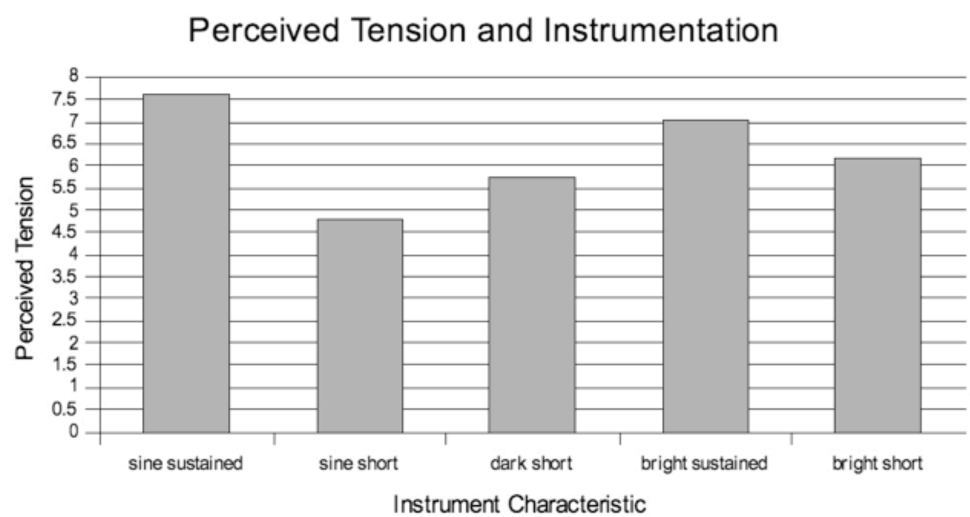

Figure 16. Geometric mean response in perceived tension as compared to change in instrumentation.

Two of the instruments were simple sine tones at the frequency of the pitch to be generated. One of the two, labeled sine sustained, had 100-millisecond overlap between notes. The other, sine short, consisted of 100-millisecond silence between each contiguous note. Comparing these sine tone instruments, we found that sustaining notes are perceived as sounding tenser than shorter resonating notes. We hypothesize that as the sustained notes overlap succeeding notes they may cause beating between the notes and therefore more distinct sensory dissonance. Additionally, we found that brighter instruments, both seen on the right in Figure 16, appeared tenser than darker sounding instruments. This finding is supported by existing research in sensory dissonance, with brighter sounds having more/stronger high frequency harmonics beating against each other (Helmholz 1885; Plomp 1964; Hutchinson and Knopoff 1978; Vassilakis 2007). In our sonification application, each fish species maps to a different musical instrument. For instance, we represent Yellow Tang with rich string sounds and the smaller Blue Chromis with a bright glockenspiel sounds. I aim for the system to eventually normalize the model of tension across instrumentation. That is, given the effect of instrumentation on perceived tension (shown in figure 16), the system should correct for this impact in order to model the same input level of tension across different instruments.

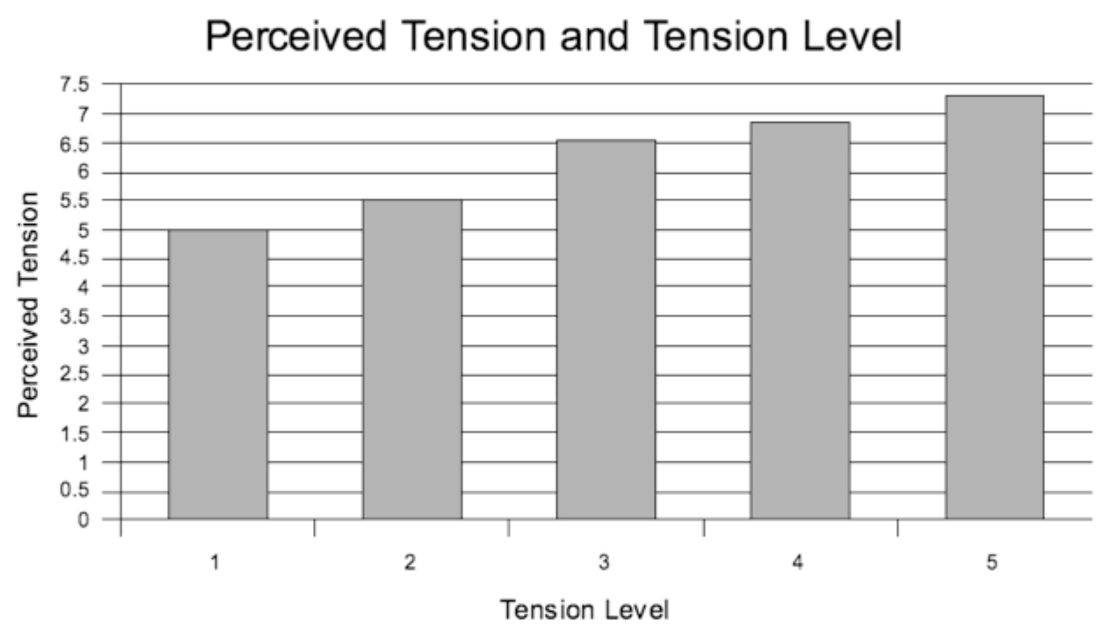

Figure 17. Geometric mean response in perceived tension as compared to change in the input tension parameter.

In evaluation of the tension control of the algorithm we compared perceived tension to the tension input level, across all manipulated conditions. Figure 17 shows the result of this analysis; with a direct linearly proportionate correlation ($r = 0.98$) between the input tension level and subjective perceived tension. This correlation demonstrates a 1:1 relationship between the tension control of our generative system and the perceived tension. It also supports the melodic tension precepts laid out by Fred Lerdahl in Tonal Pitch Space, and the effectiveness of our modifications of Lerdahl's formulas.

7.2 Study 2: Sonification

The second study evaluated the application of the system to sonification. In doing so, it further tested the claim that we can more effectively sonify data by mapping its high- and low-level features to intuitively corresponding high- and low-level musical features. A group of thirty subjects participated in this online survey in which they viewed and rated several video clips in response to a series of statements.

7.2.1 Materials

Participants viewed a series of six clips, each thirty seconds long. The clips all presented a view of an aquarium accompanied by music. Two of the clips showed the system described in this paper, in which the tracked fish positions were sonified by the generative system (one with and one without the harmonic tension model). Three others came from earlier sonification models used in the Accessible Aquarium project, as described in section 3. Finally, as a baseline, a 'de-synchronized' clip presented the video and audio from my system with the audio shifted in time, and therefore decoupling the sonification that originally took place.

7.2.2 Procedure

After viewing each of the clips participants rated their experience across several criteria. The criteria aimed to identify how effectively the music sonified the video and whether or not it improved the overall experience. The study presented the criteria as statements with which participants indicated how strongly they agreed or disagreed across a seven-point Likert scale. These statements included:

1. “The music enhances the experience of watching the video.”
2. “The music captures the most important aspects of the aquarium.”
3. “Watching this excerpt was enjoyable.”
4. “The music clearly represents what happens in the video.”

Questions 1 and 3 aim to subjectively measure the influence of music on the overall experience of watching the clips while questions 2 and 4 target more directly assess the efficacy of the sonification.

7.2.3 Results

Scoring the responses, each of the categories was analyzed independently, not summed.

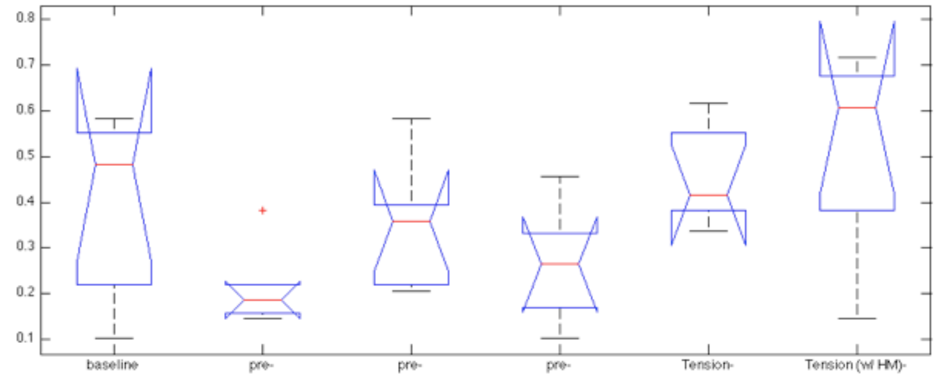

Figure 18. ANOVA analysis of responses to statement 1 "The music enhances the experience of watching the video." Baseline refers to the clip generated by my system in which the audio (originally generated by the fish movements) is offset from the video. The pre- categories refer to clips of previous implementations of the Accessible Aquarium project. The Tension category refers to video with audio generated by my system (without the harmonic expectation model) and Tension (w/ HM) refers to the same system (with the harmonic expectation model)

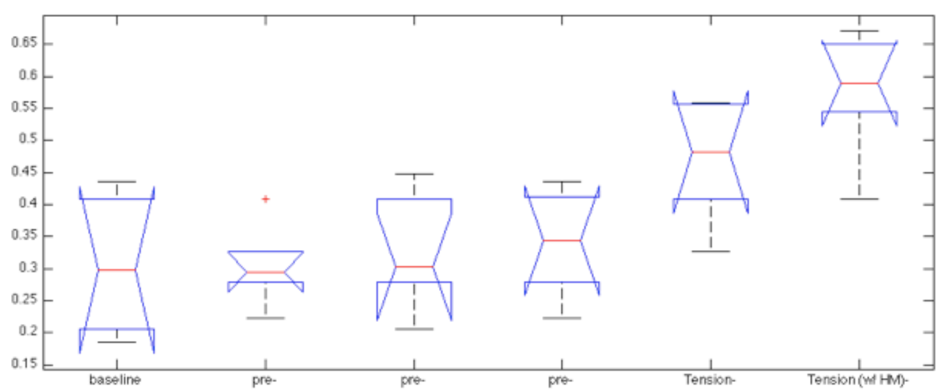

Figure 19. ANOVA analysis of responses to statement 2 "The music captures the most important aspects of the aquarium."

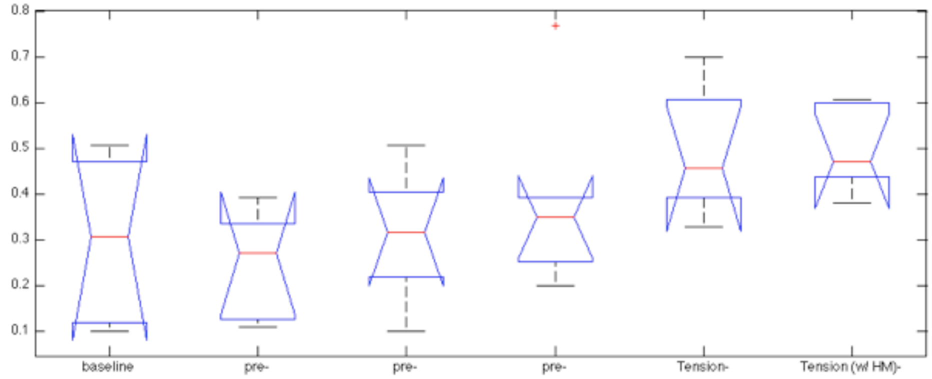

Figure 20. ANOVA analysis of responses to statement 3"Watching this excerpt was enjoyable."

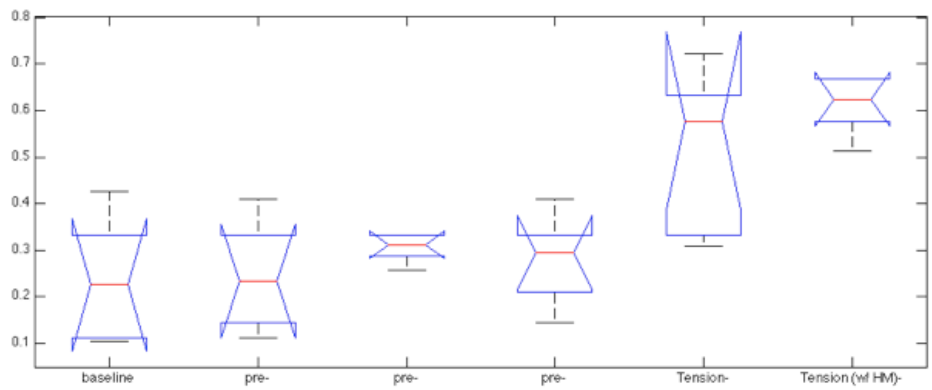

Figure 21. ANOVA analysis of responses to statement 4 "The music clearly represents what happens in the video."

While the responses to statements 1 and 3 (figures 18 and 20) seem to indicate a general preference towards the accompanying music generated by the tension system, ANOVA analysis of the results compared to the baseline did not prove statistically significant. Note, post-hoc Tukey-Kramer correction ($\alpha = .05$) was used to evaluate the significance of all of the results. Assessing the subjective response to the system's sonification task, however, participants found the tension outperformed both the baseline and previous implementations. ANOVA analysis of this data, responses to statements 2 and 4 (figures 19 and 21), showed statistically significant results ($p < .0001$) supporting the effectiveness of the tension system. Furthermore, though not statistically significant, the clip of the system with the harmonic tension model tended to outperform the implementation without the model, suggesting an additive effect in combining the models.

CHAPTER 8

CONCLUSIONS

This thesis has presented the design and implementation of a real-time music generation algorithm. As proposed, the design of this generative system modifies, builds on, and combines multiple models for analyzing tension in music. Furthermore, the thesis has contextualized the generative system with its application in sonification. Mapping features from a non-musical to musical domain, I have applied the system in sonifying the dynamic nature of aquarium activity, in particular realtime fish movements. Finally, the thesis evaluated the work through two subject-response studies. The first study supported the effectiveness of the generative system to create perceived musical tension. In the second study, while not significantly so, provided results that preferred the postulate that sonification with my system better improved the experience of viewing the aquarium. The study also provided significant results supporting my claim that the system effectively sonifies the fish movements in the aquarium.

CHAPTER 9

FUTURE WORK

While the current model described in this paper successfully addressed our originally intended goals, this work only lays the foundation for future work. In future work I hope to extend the concept of musical roles, varying degrees of leading and supportive roles, to our generative system. Finally, I want to adapt the algorithm to compensate for relative changes in tension based on information gathered from our study.

Through combinatorial processing of control parameters I also hope to explore further the full range of the system's possible generative outputs. From this study I will define distinct characteristics of the music output, provided given input parameters. I can classify these characteristics as certain musical roles. For instance, parameters limiting movement only to leaps between chord tones would most likely yield a supportive role, while increasing the likelihood of stepwise movement and non-harmonic tones may result in a more melodic and prominent lead role. Extending this to sonification, I may orchestrate the musical output, as salient moving objects, brightly colored fish, will be assigned melodic lead roles and less prominent objects, less noticeable fish, are assigned background roles of harmonic support. Additionally, the current system lacks attention to contrapuntal coherency. Future implementations will discretely address this, possibly with the superimposition of some dynamic set of contrapuntal heuristics.

In our user study I found a positive correlation from register and instrumentation to perceived tension. Based on this data, I can adjust our current model to compensate for variations in instrumentation and register. This will provide a controlled method for manipulating musical tension across varying features.

Examples of study stimuli and aquarium sonification videos can be found here: http://gtcmt.coa.gatech.edu/tension_examples/ and ryan.nikolaidis.us/Test/

APPENDIX A

CHORD TRANSITION TABLE

This table provides the probability of any chord (row) moving to any other chord (column). The (row and column) numbers represent the chord's root position above tonic and the chord's quality; the numbers are in groups of three according to quality: Major, Minor, and Diminished qualities of each pitch class root.

The number % 3 represents chord quality and number / 3 represent root position above tonic.

e.g. 0 = the Major Tonic, 3 = the Diminished Tonic, 22 = Minor Dominant

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
0	0.28	0.11	0.01	0.09	0.01	0.02	0.18	0.46	0	0.06	0.06	0.01	0.09	0.17	0.01	1	0.09	0.03	0.01	0.09	0	0.46
1	0.04	0.16	0	0	0	0	0.08	0.12	0	0.04	0	0	0	0	0	1	0.04	0	0.04	0	0	0.04
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0.18	0	0.03	0	0	0	0.32	0	0	0	0	0	0	0	0.09	0	0	0.18	0	0	0.03
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
5	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0.22	0	0	0.22	0.04	0	0.14	1	0	0.16	0.02	0	0.04	0.06	0	0	0.1	0.02	0	0.02	0	1
7	0.04	0	0	0.05	0	0	0	0.05	0	0	0.01	0.03	0.03	0.06	0	0.01	0.02	0	0	0.01	0	1
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0.11	0.06	0	0.2	0	0	0.49	0.34	0	0.03	0.2	0	0.06	0	0.03	0.03	0.03	0	0	0	0.03	0.06
10	0	0	0	0.05	0	0	0	0.32	0	0	0	0	0	0.09	0	0	0	0	0	0	0	0
11	0.43	0	0	0	0	0	0.14	0.86	0	0	0	0	0	1	0	0	0	0	0	0	0	0
12	0.02	0	0	0	0.06	0	0.06	0.04	0	0.08	0	0	0.02	0.1	0	0.15	0.02	0.02	0	0.02	0	0.04
13	0.01	0.01	0	0	0	0	0.05	0.05	0	0.05	0.05	0.05	0	0	0	0.08	0.02	0	0	0.01	0	0.02
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
15	1	0.05	0	0.03	0	0	0.3	0.24	0	0.05	0	0	0.24	0.19	0	0.35	0.57	0	0.08	0.22	0.43	0.49
16	0.25	0.03	0	0	0	0	0	0.03	0	0.06	0.03	0	0.06	0.36	0	0	0.06	0	0	0	0	0.11
17	0	0	0	0	0	0	0.25	0	0	0	0	0	0	0	0	1	0	0	0	0.25	0	0
18	0.11	0	0	0.11	0.11	0	0	0	0	0.11	0	0	0	0.11	0	1	0.33	0	0	0.22	0	0.11
19	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03	0.17	0	0.03	0	0	0.06
20	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0	0	0	0
21	1	0.05	0.01	0	0	0	0.03	0.1	0	0.01	0	0.01	0.02	0.12	0	0.03	0.01	0	0.02	0.01	0	0.07
22	1	0.02	0	0	0	0	0	0	0	0.05	0	0	0	0	0	0.02	0	0	0.08	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0
24	0.12	0.05	0	0.32	0.07	0	0	0.93	0	0.07	0.02	0	0	0.05	0	0.07	0.1	0	0	0	0	1
25	0	0	0	1	0	0	0.33	0.33	0	0	0	0	0	0	0	0	0.33	0	0	0	0	0.67
26	0	0	0	0	0	0	0.4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0.05	0	0	0.02	0	0	0.41	1	0	0	0.01	0	0	0.06	0	0.02	0.02	0	0.01	0.04	0	0.01
28	0	0	0.02	0	0	0	1	1	0	0.14	0.02	0	0.1	0.02	0	0.07	0.02	0	0	0.07	0	0.12
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0.71	0	0	0.04	0	0	0.04	0.21	0	1	0.14	0	0.11	0.46	0	0	0.07	0	0	0.07	0	0.18
31	0.04	0	0	0	0	0	0	0	0	1	0.04	0	0.28	0	0	0.04	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0.89	0	0	0.11	0	0	0.05	0.05	0	0	0.05	0	1	1	0	0	0.11	0	0	0.11	0	0.16
34	0.03	0.03	0	0	0	0	0.03	0.03	0	0	0	0	1	0	0	0.05	0	0	0	0	0	0
35	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	22	23	24	25	26	27	28	29	30	31	32	33	34	35
0	0.14	0	0.11	0.01	0	0.36	0.2	0	0.14	0.04	0	0.06	0.07	0
1	0.12	0	0	0	0	0	0	0	0.08	0.08	0	0.04	0.16	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0.33
3	0.03	0	0.03	0	0	0	0	0	0	0	0	0.09	0	0
4	0	0	0.33	0	0	0.33	0	0	0	0	0	1	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0.1	0	0.1	0.06	0.08	0.02	0.06	0	0	0	0	0.02	0	0
7	0.01	0	0.01	0	0.01	0.01	0	0	0.03	0	0	0	0.01	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	1	0	0	0.03	0.11	0	0.03	0	0	0	0	0
10	0	0	1	0	0	0.05	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0.15	0	0	0.52	1	0	0.02	0.1	0	0	0	0
13	0.01	0	0.04	0	0	1	0.25	0	0	0	0	0	0.01	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0.27	0	0.14	0	0	0.03	0.08	0	0.73	0.38	0	0.08	0.43	0
16	0.03	0	0	0	0	0	0	0	1	0.06	0	0.06	0.03	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0.11	0	0	0	0	0	0.56	0	0
19	0	0.03	0	0	0	0.03	0	0	0	0	0	1	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0
21	0.05	0	0.04	0	0	0.01	0.02	0	0.02	0	0	0.01	0	0
22	0.05	0	0	0	0	0	0.05	0	0	0	0	0.02	0.03	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	0.12	0	0	0.02	0	0.2	0.02	0	0.02	0	0	0.12	0	0
25	0	0.33	0	0.33	0	0	0	0	0	0	0	0	0	0
26	1	0	0	0	0	0	0.2	0	0	0	0	0	0	0
27	0.02	0	0.09	0	0	0.04	0.07	0	0.08	0	0	0	0.04	0
28	0.1	0	0.1	0.05	0.02	0.1	0.38	0	0.05	0.05	0	0	0.17	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0.04	0	0	0.54	0.11	0	0.07	0.43	0	0.32	0	0
31	0	0	0.44	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0.05	0.11	0	0	0	0	0.53	0.05	0	0.21	0.11	0
34	0	0	0	0	0	0.03	0.05	0	0	0.08	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0	0

REFERENCES

Belzer, Jack. Holzman, Albert. Kent, Allen. Encyclopedia of computer science and technology. 11th vol. 1981. pp. 80-81.

Ben-Tal, O. and Berger, J. Creative aspects of sonification, Leonardo, 20 (3), MIT Press 2004, 229-232.

Berthoz, A. 2000. The brain's sense of movement. Cambridge, MA: Harvard University Press.

Bharucha, J.J. 1984. Anchoring effects in music: The resolution of dissonance. Cognitive Psychology, 16, 485-518.

Bharucha, J. 1987. Music cognition and perceptual facilitation: A connectionist framework.a Music Perception, 5(1).

Bharucha, J. 1993. "MUSACT: A Connectionist Model of Musical Harmony." In Machine Models of Music, eds. S. M. Schwanauer and D. A. Levitt, Cambridge, Mass.: MIT Press, pp. 497-509.

Bharucha, J.J. 1996. Melodic anchoring. Music Perception, 13, 383-400.

Bharucha, J. J., and Stoeckig, K. 1986. Reaction time and musical expectancy: Priming of chords. Journal of Experimental Psychology: Human Perception and Performance, 12, 403–410.

Bharucha, J. J. and Stoeckig, K. 1987. Priming of chords: Spreading activation or overlapping frequency spectra? Perception & Psychophysics, 41:519–524.

Boll, S., Heuten, W., Wichmann, D. 2006. "Interactive 3D sonification for the exploration of city maps". In Nordic Conference on Human-Computer Interaction; Vol. 189. pp. 155-164.

Brewster, S., Ramloll, R., Yu, W., 2000. "Constructing Sonified Haptic Line Graphs for the Blind Student: First Steps." In Proceedings of ACM Assets 2000, ACM Press, pp. 17-25.

Butler, D. A study of the event hierarchies in tonal and post-tonal music. Psychology of Music, Vol. 18, No. 1, 4-17, 1990.

Cope, D. 1991. "Computer Simulations of Musical Style." In Computers in Music Research, The Queens University of Belfast, 7-10 (April), pp. 15-17.

Cope, D. David Cope. 6 June 2009. <http://arts.ucsc.edu/faculty/cope/experiments.htm>.

Davison, B., and Walker, B. N. 2007. Sonification Sandbox overhaul: Software standard for auditory graphs. Proceedings of the International Conference on Auditory Display (ICAD 2007), Montreal, Canada (26-29 June). pp. 509-512.

Deliège, I. 1997. Perception and Cognition of Music. Hove, East Sussex Psychology Press. pp. 334.

Desain, P. A (de)composable theory of rhythm perception. Music Perception, 9(4), 439-454, 1992.

Desain, P. and Honing, H. 2001. Modeling the effect of meter in rhythmic categorization: Preliminary results. Japanese Jour- nal of Music Perception and Cognition, 7(2):145–56.

Desain, P., and Honing, H. 2002. “Rhythmic Stability as Explanation of Category Size.” Paper presented at the 2002 International Conference on Music Perception and Cognition, University of New South Wales, Sydney, July 17–21.

Deutsch, D. 1986. Recognition of durations embedded in temporal patterns. Perception & Psycho-physics, 39, 179–187.

Dodge, C. 1962. Earth's Magnetic Field. Nonesuch.

Drobisch, M. W. 1855. Über musikalische Tonbestimmung und Temperatur“. (Abhandlungen
der Mathematisch-Physischen Klasse der Sächsischen Akademie der Wissenschaften, 2).

Dubois, L. harddata. 7 June 2009. <http://www.turbulence.org/Works/harddata/>.

Dunn, J. and Clark, M. 1999. "Life Music": The Sonification of Proteins. Leonardo, Vol. 32, No. 1, pp. 25-32

Engel, A.K., Fries, P., Singer,W. 2001. Dynamic predictions: oscillations and synchrony in top–down
processing, Nature Reviews Neuroscience 2, pp. 704–16.

Farbood, M. 2001. A Quantitative, Parametric Model of Musical Tension. Diss. Massachusetts Institute of Technology. 9 June 2009. <http://web.media.mit.edu/~mary/>.

Flowers, J., Whitwer, L., Grafel, D. and Kotan, C. 2001 "Sonification of daily weather records: Issues of perception, attention and memory in design choices," presented at International Conference on Auditory Display, Espoo, Finland.

Franklin, J. 2001. Learning and Improvisation, Neural Information Processing Systems 14, eds. Dietterich, T. G.,Becker, S., & Ghahramani, Z. Cambridge, MA: MIT Press.

Fraisse, P. 1982. Rhythm and Tempo. The Psychology of Music. pp.149-180. New York: Academic Press.

Hagendoorn I (2004) Some speculative hypotheses about the nature and perception of dance and choreography. J Consci St 11:79–110

Helmholtz, H. 1885 [1954]. On the Sensations of Tone as a Physiological Basis for the Theory of Music. 2nd English edition. New York: Dover Publications. [Die Lehre von den Tonempfindungen, 1877. 4th German edition, trans. A. J. Ellis.]

Hermann, T. et al. 2002. "Sonifications for EEG Data Analysis." Proc. Int'l Conf. Auditory Display, Int'l Community for Auditory Display. pp. 37-41.

Huron, D. 2001. Tone and voice: A derivation of the rules of voice-leading from perceptual principles. Music Perception, 19, 1–64.

Hutchinson, W. and Knopoff, L. 1978. “The acoustic component of Western consonance,” Interface, vol. 7, pp. 1-29.

Jonaitis, E. Saffran, J. 2009. Learning the Role of Serial Statistics. Cognitive Science. 951-968.

Justus, T. Bharucha, J. 2001. Modularity in Musical Processing: The Automaticity of Harmonic Priming. J. Experimental Psychology: Human Perception and Performance. Vol. 27, No. 4, 1000-1011.

Keller, D. Ferneyhough, B. Analysis by Modeling: Xenakis's ST/10-1 080262. Journal of New Music Research, Volume 33, Number 2, June 2004 , pp. 161-171(11).

Kohonen, T. 1989. A self-learning musical grammar, or "Associative memory of the second kind". In Proceedings of the International Joint Conference on Neural Networks, New York: IEEE, pp. 167-186.

Kramer, G., Walker, B. N., Bonebright, T., Cook, P., Flowers, J.H., Miner, N., Neuhoff, J., et al. 1997. Sonification report: status of the field and research agenda (1997). http://www.icad.org/websiteV2.0/References/nsf.html.

Krumhansl, C.L. 1979. The psychological representation of musical pitch in a tonal context.
Cognitive Psychology, 11, 346-384.

Krumhansl, C. L., and Kessler, E. J Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychological Reuiew, 1982,89, 334–368.

Krumhansl, C.L., Toivanen, P., Eerola, T., Toivianinen, P., Järvinen, T., & Louhivuori, J. .2000. Cross-cultural music cognition: Cognitive methodology applied to North Sami yoiks. Cognition, 76, 13-58.

Larson, S. 1994. Musical forces, step collections, tonal pitch space, and melodic expectation.
Proceedings of the Third International Conference on Music Perception and Cognition, 227-229.

Larson, S. 1997. The problem of prolongation in tonal music: Terminology, perception, and expressive meaning. Journal of Music Theory, 41, 101–136.

Larson, S. 2004. Musical forces and melodic expectations: Comparing computer models and
experimental results. Music Perception, 21, 457-98.

Lerdahl, F. and Jackendoff, R. A Generative Theory of Tonal Music. The MIT Press, 1996.

Lerdahl, F. and Krumhansl, C. 2007. Modeling Tonal Tension. Music Perception, 24, 329-366.

Lerdahl, F. 2001. Tonal Pitch Space. Oxford University Press, New York.

Margulis, E. 2005. A Model of Melodic Expectation. Music Perception, 22(4):663–714.

Meyer, L.B., 1956. Emotion and Meaning in Music. University of Chicago Press, Chicago.

Moorer, J. 1993. "Music and Computer Composition." In Machine Models of Music, eds. S. M. Schwanauer and D. A. Levitt, Cambridge, Mass.: MIT Press, pp. 167-186.

Narmour, E. 1992. The Analysis and Cognition of Melodic Complexity: The Implication-Realization Model. Chicago: University of Chicago Press.

Pachet, F. 2002. Playing with Virtual Musicians: the Continuator in practice, IEEE Multimedia 9:3, pp. 77-82.

Pachet, F. 1999 Surprising Harmonies. JIM 99, 6e`mes journe´es d'informatique musicale, CNET-CEMAMu, Issy-les-Moulineaux, pp. 187–206

Paine, G. (2004). Reeds – a responsive sound installation. Proc. of the International Conference on Auditory Display (ICAD 2004)

Pendse, A. Pate, M. and Walker, B. 2008. "The Accessible Aquarium: Identifying and Evaluating Salient Creature Features for Sonification". ASSETS'08, October 13–15, 2008, Halifax, Nova Scotia, Canada.

Plomp, R.1964."The ear as a frequency analyzer," J. Acoust. Soc. Am., vol. 36(9), pp. 1628-1636.

Polli, A. 2004. Atmospherics/weather works: A spatialized meteorological data sonification project. In International Conference on Auditory Displays (ICAD), pp. 31–36.

Povel D-J. 1981. Internal representation of simple temporal patterns. J. Exp. Psychol.: Hum. Percept. Perform. 7:3–18

Rubin, B. and Hansen, M. Listening Post - EAR Studio. May, 8, 2010. http://earstudio.com/2010/09/29/listening-post/

Shepard, R. N. 1982. Geometrical approximations to the structure of musical pitch. Psychological Review 89 (1982), 305-333.

Steinbeis, N. Koelsch, S. and Sloboda, J. 2006. The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. J. Cogn. Neurosci., 18(8):1380-1393.

Stevens, S. S. 1975. Psychophysics: Introduction to its Perceptual, Neural, and Social Prospects. Wiley, New York.

Tamm, E. 1995. Brian Eno: His music and the vertical color of sound. New York: Da Capo.

Thom, B. 2001. BoB: An improvisational music companion. PhD dissertation, School of Computer Science, Carnegie-Mellon University.

Tillmann, B., Janata, P., Birk, J., & Bharucha, J. J. 2003. The costs and benefits of tonal centers for chord processing. Journal of Experimental Psychology: Human Perception and Performance, 29, 470–482.

Tillmann, B., Janata, P., Birk, J., & Bharucha, J. J. 2008. Tonal centers and expectancy: Facilitation or inhibition of chords at the top of the harmonic hierarchy? Journal of Experimental Psychology. Human Perception and Performance, 34, 1031–1043.

Vassilakis, P. and Fitz, K. 2007. SRA: A Web-based Research Tool for Spectral and Roughness Analysis of Sound Signals. Supported by a Northwest Academic Computing Consortium grant to J. Middleton, Eastern Washington University.

Walker, B. 2002. Magnitude estimation of conceptual data dimensions for use in sonification. Journal of Experiment Psychology: Applied, 8.

Walker, B., Kim, J., and Pendse, A. 2007. Musical soundscapes for an accessible aquarium: Bringing dynamic exhibits to the visually impaired. Proceedings of the

International Computer Music Conference (ICMC 2007), Copenhagen, Denmark (27-30 August). pp. 268-275.

Weinberg, G., and Thatcher, T. 2006. Interactive sonification of neural activity. In NIME '06: Proceedings of the 2006 conference on New interfaces for musical expression, Paris, France, 2006. IRCAM. pp. 246– 249

Wilson, J., Walker, B. N., Lindsay, J., Cambias, C., & Dellaert, F. 2007. SWAN: System for Wearable Audio Navigation. Proceedings of the 11th International Symposium on Wearable Computers (ISWC 2007), Boston, MA (11-13 October).

Zicarelli, D. 1988. "M and Jam Factory," Computer Music Journal 11, No. 4.